Behavioral Finance and Swiss Pension Plans
A Collection of three Essays

Dissertation
for the Faculty of Economics, Business Administration
and Information Technology of the University of Zurich

to achieve the title of
Doctor of Economics

presented by
Christoph Gort
from Pfäfers (Switzerland)

approved at the request of

Prof. Dr. Thorsten Hens
Prof. Dr. Mei Wang

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich herewith permits the publication of the aforementioned dissertation without expressing any opinion on its views.

Zurich, December 2007

The Dean: Prof. Dr. H.P. Wehrli

For Marcella and Peter

Acknowledgements

This PhD Dissertation is the result of my research at the University of Zurich. I would never have completed this work without the help and support of many people, to whom I owe my thanks.

First I want to thank my thesis adviser Prof. Dr. Thorsten Hens for his guidance and coaching. On several occasions he gave me helpful advice and motivated me to continue my research efforts. Many thanks also go to my coauthors, Prof. Dr. Mei Wang, Prof. Dr. Michael Siegrist, and Prof. Dr. Anke Gerber who provided important methodological and substantive advice. They contributed greatly to the quality of my thesis.

In addition to my coauthors I would like to thank the participants of the "Schwerpunkt Finance" meetings who contributed not only on the methodological level but also with the interpretation of the results of my research efforts. Special thanks go to Dr. Ulf-Dietrich Reips, Prof. Dr. Klaus Jonas, Dr. Alexander Wagner, Prof. Dr. Janos Mayer, Prof. Dr. Enrico de Giorgi, Dr. Marc-Oliver Rieger, Prof. Dr. Marc Chesney, Martin Vlcek, Kremena Bachmann and all the other people who made valuable comments. I thank Andreas Tupak and Sonja Näf for providing technical support in programming the Z-Tree code and helping with my market experiments.

In my personal environment Dr. Vera Kupper, Dr. Jürg Tobler, Dr. Christian Fitze, Dr. Thomas Häfliger, Dr. Urban Müller, Dr. Mark Schindler, Andrew Marks, Leslie Hill, Nigel Cresswell, Geraldine Haldi and Dr. Andreas Reichlin gave me lots of valuable inputs. I also thank ASIP, Watson Wyatt and PPC Metrics for their technical support and for the provision of empirical data.

Financial Support from the members of the University Priority Program "Finance and Financial Market" at the University of Zurich is gratefully acknowledged.

Last but not least I thank my parents, Marcella and Peter, for the wonderful support I have enjoyed all years.

Contents

List of Figures

List of Tables

1. Introduction

My PhD Dissertation consists of 3 independent essays. One thing all three essays have in common is that they focus on financial decision-making. During my practical experience at the City of Zurich Pension Fund I frequently observed various different sorts of financial decision-making processes and those observations caught my attention. First, because several factors might bias the perception, the importance and the rationality of a financial decision and second because the characteristics of the decision-maker(s) might matter. This first chapter mainly serves as an introduction into the topics of my Dissertation and outlines the reasons why I have chosen this topic in general.

The first essay deals with individual estimations of returns on financial markets in a sample of Swiss pension plan managers. The second one asks Swiss pension plan managers about the abilities of their own pension plan to achieve above average returns on financial markets. The main goal of those two essays is to analyze the degree of overconfidence among the participants based on a self-developed questionnaire. The third essay addresses differences between groups and individuals in financial decision-making on an experimental market. Each essay will in detail address their academic as well as their practical implications.

I realized that many people, especially in the asset management industry, are very confident about their knowledge and their abilities to beat market index returns. Most active asset managers usually claim to outperform the market in the future. Most of the pension funds in Switzerland manage their assets actively and try to outperform market indices. However, beating a broadly diversified market-capitalized stock index is very difficult as Sharpe (1991) has demonstrated. Most of the academic studies report that only a minor percentage of investors realize an outperformance in the long run after costs (Carhart (1997) or Malkiel (2004)). So the basic goals of my first two essays were to test and to describe the observed confidence levels in a scientific way and to relate those findings to the way pension plans manage their assets today.

One area of today's behavioral finance literature deals with the phenomenon of overconfidence which offers a perfect starting point to address those basic goals. Overconfi-

dence is a complex phenomenon with various facets including miscalibration and the better-than-average-effect (Glaser and Langer (2003)). Miscalibration refers to the fact that people provide very narrow confidence intervals in various estimation tasks (Lichtenstein, Fischhoff and Philips (1982)). The better-than-average-effect describes the evidence that most people believe to achieve above average performances in various fields (Taylor and Brown (1988)). Both of these psychological findings can also be applied in the domain of financial markets. Therefore, I dedicated my research to those two aspects of overconfidence by analyzing a set of Swiss pension plan managers in an empirical study. A sample of Swiss pension plan managers is distinct from previous studies about private investors' or mutual fund managers' degrees of overconfidence and their realized performances with active management because the participants in my sample not only bear responsibility for their own investments but for the retirement savings of all the employees in Switzerland. Therefore an additional level of prudence from those participants could be expected. In contrast Odean (1998) outlines that possibly exactly those people who are overconfident in the domain of financial markets are those who are attracted by jobs that require financial decision-making. The main results of my first two essays show that Swiss pension plan managers are indeed overconfident. They provide very narrow confidence intervals when estimating returns on financial markets and most of them believe the active managers of their own pension plan to be above average.

As previous research shows overconfidence can lead to reduced returns, increased risks and suboptimal financial decision-making. So it seems very important to address those issues in order to help investors, especially Swiss pension plans, to manage their assets more efficiently.

Another important observation during my time with the City of Zurich Pension Fund is that investment decisions can be taken in very different ways. The investment committee of the City of Zurich Pension Fund usually discusses various ideas internally and then takes a decision. This is a typical example of a group decision-making process. In contrast some of the external asset managers the City of Zurich Pension Fund has employed rely on the decision-making of an individual, usually a senior portfolio manager. Today, neither psychological nor economic research offer a clear answer to the question whether groups outperform individuals in financial decision-making and therefore I

focused on this question in my third essay. The lack of an answer motivated me to design an asset allocation game on an experimental market which allows me to observe the decision-making process of groups and individuals in a sample of students and to test whether there are any significant differences between groups' and individuals' performances.

The main conclusion is that only groups who exchange individual information across group members are able to achieve better results than individuals. This has important implications for the decision-making processes of asset managers on financial markets because they should take the results of my third essay into account when organizing their financial decision-making structures.

The following chapters 2, 3, and 4 contain the full version of each of my 3 essays including the abstract, all the figures and tables, and the corresponding references.

2. Are Pension Fund Managers Overconfident?

Abstract

Empirical studies show that people tend to be overconfident about the precision of their knowledge, leading to miscalibration. Consistent with this, we found that on overage the decision makers of Swiss pension plans provide too narrow confidence intervals when asked to estimate past returns of various assets. Their confidence intervals are also very narrow in their forecasts of future returns. They are less miscalibrated, however, than our laypeople sample. Individual differences between the participants' degree of overconfidence are large and stable across those two different tasks. In a linear regression model we present evidence that the size of participants' confidence intervals is linked to individual characteristics. In our sample younger people with a degree from university and with more experience in finance provide larger intervals than older people without such an education and with less experience.

2.1 Introduction

On average people tend to be overconfident. In particular, it is well documented that people exhibit overconfident behavior in financial markets. The degree of overconfidence, however, seems to vary across individuals and across different domains of questions. In this paper our contribution to research is twofold. First, we investigate a special group, the decision-makers of Swiss pension plans who not only bear responsibility for their own investments but for the retirement savings of thousands of employees in Switzerland. Therefore an additional level of prudence from those participants could be expected. In contrast Odean (1998) outlines that possibly exactly those people who are overconfident in the domain of financial markets are those who are attracted by jobs that require financial decision-making. We shed some light on this question by showing that decision-makers of Swiss pension plans are overconfident but to a lesser degree than a sample of laypeople.

Second, we not only confirm the evidence for individual differences in the degree of overconfidence but also show that those differences are related to individual characteristics. In a linear regression model we measure the impact of individual characteristics on overconfidence and we present evidence that younger people with a degree from university and more experience in finance are significantly less overconfident than older participants with less education and less experience.

The remainder of the paper is organized as follows. Section II reviews related research on overconfidence in general and in the domain of financial markets. Section III describes the data and the methods to measure overconfidence and introduces our linear regression model. Section IV presents the results for miscalibration in our sample and for our linear regression analysis. Section V discusses interpretations and practical implications of our results and concludes.

2.2 Related research on overconfidence

Overconfidence is a complex phenomenon with various facets. Glaser and Weber (2003) differentiate between four different manifestations of overconfidence: miscalibration, better-than-average-effect, illusion of control and overoptimism. This paper concentrates only on i) miscalibration in the domain of estimating historical returns of financial assets and on ii) forecasting future returns.

People tend to overestimate the precision of their knowledge. As a result, they are miscalibrated in estimating and forecasting by providing too narrow confidence intervals (Lichtenstein, Fischhoff and Philips (1982)). It has been observed that task difficulty and blurred feedback lead to more overconfidence (Lichtenstein, Fischhoff and Philips (1982), Griffin and Tversky (1992)). Odean (1998) argues that forecasting and estimating returns on financial markets are not easy tasks and the available feedback is blurred as the market prices of assets are affected by noise. So the chances to observe overconfident behavior in the domain of financial markets are high.

Current psychological research debates whether miscalibration is a stable human trait or only a statistical illusion (see Gigerenzer, Hoffrage and Kleinbölting (1991), Griffin and Tversky (1992), Erev, Wallsten and Budescu (1994), Brenner, Liberman and Tversky (1996) and Klayman, Soll, Gonzales-Vallejo and Barlas (1999)). As Soll and Klayman (2004) point out the type of question matters and tasks which involve estimations of confidence intervals typically lead to higher measures for miscalibration. It is beyond the scope of this paper to analyze miscalibration in general and with respect to different types of measurement. We focus on the tasks of estimation and forecasting of asset returns, which are similar to tasks the decision makers of Swiss pension plans frequently face in their jobs and which might impact the wealth of Swiss pension plans.

Studies show that professionals in the financial industry are subject to miscalibration. Russo and Schoemaker (1992) report that money managers tend to formulate too narrow 90% confidence intervals in a questionnaire about meta-knowledge. The participants' subjective confidence intervals in their sample contain the correct solutions only in about half of the cases instead of 90% as required. Graham and Campbell (2003) analyze economic forecasts on the equity risk premium from CFOs in the USA over different time horizons and conclude that the size of the average confidence interval is very narrow compared to the volatility of equity markets. Deaves, Lueders and Schroeder (2005) and Glaser, Weber and Langer (2003) present similar evidence in the domain of financial markets as the confidence intervals of the participants in their samples of professionals capture significantly less realized returns for economic forecasts than required. They also notice that the individual degree of overconfidence is stable across different tasks. This result indicates that people are in general overconfident in the domain of financial markets and not just within particular asset classes or particular tasks.

It is also in line with the finding of Alpert and Raiffa (1982) who argue that people tend to respond similarly to the same types of questions.

Graham, Campbell and Huang (2006) and Glaser, Weber and Langer (2005) report that the level of overconfidence in the domain of financial markets is different across individuals. There is no doubt that individual characteristics affect overconfidence but the evidence about stable relationships is ambiguous. Russo and Schoemaker (1992) report that professionals are generally miscalibrated but to a lesser degree than laypeople. In contrast Glaser, Weber and Langer (2003) find that professionals are more overconfident than students about their trend recognition abilities although they do not provide more accurate estimations. In a model from Odean and Gervais (2001) more experience is related to a lower degree of overconfidence. Inexperienced but successful investors are most prone to overconfidence as they self-attribute their success solely to their abilities. Over time more experience will help them to better evaluate their true abilities. Locke and Mann (2001) confirm this theory empirically as they find no indication of miscalibration among highly experienced traders on the Chicago Mercantile Exchange. In light of those results we analyze if individual characteristics such as education or experience increase overconfidence in the domain of financial markets.

Being overconfident can be harmful on financial markets. In a large sample of private investors Odean and Barber (2001) show that overconfidence leads to a higher trading volume and reduces portfolio returns. Guiso and Jappelli (2005) use a sample of Italian bank clients in which the clients, - whom the authors suppose to be more overconfident (people with a lower education but a higher self declared knowledge), - hold portfolios with lower Sharpe ratios than other clients. However it is beyond the scope of this paper to evaluate the portfolios and the trading activities of Swiss pension plans and we do not postulate any causal relationships between the degree of overconfidence of our participants and the investments of the corresponding pension plans.[1]

[1] Menkhoff, Lukas, and Ulrich Schmidt, 2005, The Use of Trading Strategies by Fund Managers: Some First Survey Evidence, Discussion Paper (Hannover). The authors address how overconfidence might be related to investment strategies of mutual fund managers.

2.3 Data and methods

In total 584 questionnaires have been distributed among decision-makers of Swiss pension plans, i.e. managers and members of investment committees and 132 have been returned. We refer to it as the professional sample. This corresponds to a response rate of 22.6%. 24 questionnaires contained no confidence intervals and therefore have been excluded from the analysis so the professional sample consists of 108 participants (Only 6 participants are female). 58 persons have a university degree, 36 of them in finance. 65 attended education courses in finance for practitioners. Experience in finance and in pension plans is symmetrically distributed between less than 2 years and more than 25 years, and the respondents are between 25 and 80 years old. A laypeople sample is based on people working for the City of Zurich in several departments not related to financial markets or pension plans but with a self-declared interest in financial topics. In total 104 persons, 19 woman and 85 men, returned a complete questionnaire so the sample size of the laypeople sample is 104 persons (22 questionnaires were incomplete). 25 of them have a degree from university but only 16 in finance or economics and 32 have attended courses in finance for practitioners. Two-thirds have no experience in working in financial areas but two-thirds do frequently read newspapers related to financial topics. The participants in the laypeople sample are between 20 and 65 years old.

The questionnaire for the participants in both the professional and laypeople sample consists of two parts. In the first part the respondents provided data about individual characteristics. In the second part the participants were asked to formulate two sorts of 90% confidence intervals.[2] First, 90% confidence intervals for historical annual returns for 6 different asset classes over the last 36 years to estimate the participant's degree of miscalibration. Those questions have been worded the following way: "*In this task you have to provide an upper and a lower boundary for the annual returns of asset class x over the last 36 years. Please choose the boundaries in such a way that 90% of the realized single annual returns of asset class x are within your boundaries.*" Second, 90% confidence intervals for return forecasts for 6 different asset classes for the year 2006 to qualitatively assess how confident the participants are about their forecasting abilities

[2] A full version of the questionnaire -either in German or in French - can be obtained from the first author of this paper.

given their confidence intervals. The following wording is used: *"In this task you have to provide an upper and a lower boundary for the return of asset class x in the year 2006. Please choose the boundaries in such a way that the realized return of asset class x in 2006 will be within your boundaries with a probability of 90%."*

The period for handing in the questionnaire was from May 2006 until the beginning of August 2006. The returns of the different asset classes were volatile over that time period so maybe the 90% confidence intervals were affected. A t-test reveals however that there are no differences between the means for 90% confidence intervals from people who handed in their questionnaires before or after mid of June 2006 so there is no need to split our samples.

In this paper we use two different methods to judge the participant's confidence intervals. First, following an idea of Hilton (2001), we compare the participants' subjective confidence intervals for annual returns with the distribution of historical annual returns over the last 36 years. More concretely we count the number of annual returns over the last 36 years that are included within the participants' 90% confidence intervals. For each asset class we collected the last 36 realized historical annual returns and we simply cut off the 2 highest and lowest returns to approximate a 90% (precisely 88.9%) interval of the annual returns in each asset class. In other words 90% of the annual returns over the last 36 years are included in those intervals and this corresponds to 32 annual returns. A miscalibrated participant will provide too narrow 90% confidence intervals and thus captures less than 32 annual returns of the last 36 annual returns.

Second, we analyze the implied volatility of the participants' confidence intervals where we make use of a relationship that Pearson and Tukey (1965) describe. With the term implied volatility we refer to a relationship between the 95% and 5% return quantile (which corresponds to a 90% interval) and the standard deviation as is given in equation (1).

$$\text{Standard deviation} = (95\%\ \text{return quantile} - 5\%\ \text{return quantile}) / 3.25 \qquad (1)$$

Like in the first approach we then compare the participants' answers with historical data, i.e. the participants' implied volatilities in their confidence intervals with the historically implied volatility of each asset class based on the annual returns over the last 36 years. Volatility is a popular way to express uncertainty about future returns of an

asset class and the higher the volatility the broader is the spectrum in which the realization of the future return will fall with a certain probability. If a participant formulates confidence intervals with very low implied volatilities we interpret this as an indication that he is miscalibrated. The historically implied volatility of the annual returns therefore serves as a guideline to judge the size of the implied volatilities.

We do not apply those two methods to measure miscalibration in the forecasting task because an ex post comparison between a subject's forecast interval for the return of an asset class in the year 2006 and the accuracy of his answer might be biased and unreliable. The reason is that the measurement is heavily dependent on the future outcome, i.e. the realized return in 2006. The annual return in 2006 represents only one single observation which is not necessarily representative for the participant's true level of overconfidence. Returns close to the historical mean will lead to the conclusion that few participants are miscalibrated whereas extremely positive or negative returns would probably fall out of almost everybody's confidence intervals. Nevertheless forecast intervals provide information about how participants evaluate their own abilities to forecast future returns on financial markets and narrow confidence intervals reflect a high conviction.

To analyze the relationships between confidence interval sizes and individual characteristics we use a linear regression model with 4 predictors. We differentiate between 3 sorts of education: a degree from university, practical financial education, and no such education, resulting in 2 dummy variables. We further apply a predictor for experience – an aggregation of experience in finance and in pension plans – and age. Some of those 4 predictors are positively correlated but never above a level of 0.6. Gender is not included in the regression model as the number of females is too low.

In the analysis of confidence intervals across all participants we present median values as there are a few outliers that have big impacts on the mean. In the application of the linear regressions analysis we use the logarithm of the confidence intervals and the corresponding boundaries to mitigate such outlier effects. However the results do not substantially change if we analyze the non-transformed data. To aggregate confidence intervals of different asset classes we normalize the data by calculating the z scores of the participants log intervals. We report R^2 to provide information about the amount of variance our regressions explain. No significant interaction effects have been identified

within the variables for our regression model so we do not include interaction variables. Cooks Distance values (Cooks D) indicate no significant effects of outliers.

2.4 Results

2.4.1 Estimation of historical returns

The median 90% confidence intervals of professionals only capture around 60%~80% of the past annual returns in each asset class.[3] This is evidence that professionals in our sample are miscalibrated in the median because their confidence intervals were meant to contain 90% of the annual returns over the last 36 years. The laypeople sample provides even narrower boundaries for almost all asset classes so laypeople are more miscalibrated than professionals. A Mann-Whitney test however reveals that the differences between the confidence interval sizes are not significant except for 2 asset classes. The bars in figure 2.1 show the median lower and the median upper boundaries for the confidence intervals in the professional and the laypeople sample for historical return estimates in 6 different asset classes. Those are Swiss and world equities, CHF bonds, gold, oil and the CHF-USD exchange rate. The bars with the diagonal lines represent the distribution of 90% of all annual returns over the last 36 years for those 6 asset classes and serve as a guideline to judge the size of the participants' confidence intervals. It can be seen that in the median the professionals and the laypeople underestimate the downside risk but also the upside potential in almost all asset classes. That explains why medians for 90% confidence intervals from both samples are narrower than the historical 90% intervals for annual returns over the last 36 years.[4]

[3] In the analysis of confidence intervals for historical returns we only include participants who provided negative lower boundaries in the confidence intervals for the asset class world equities. The reason is to not bias the study with respondents who might have misunderstood the question (i.e. provided a 90% confidence interval for the annualized mean return over the whole 36 years instead of a 90% confidence interval for all annual returns in that period). In total 56 participants from the professional sample are included. If we included the confidence intervals from all the participants, miscalibration would appear to be much higher but arguably may be spurious. In the analysis of confidence intervals for return forecasts we include all participants.

[4] We acknowledge that an analysis of only the size of the participants' confidence intervals does not provide any information about the accuracy, i.e. how much the lower and upper boundaries of a subjective confidence interval deviate from the historical intervals. Mann-Whitney tests show that the professional sample provides significantly more accurate boundaries for every asset class than the laypeople.

Table 2-1: Implied volatilities of confidence intervals

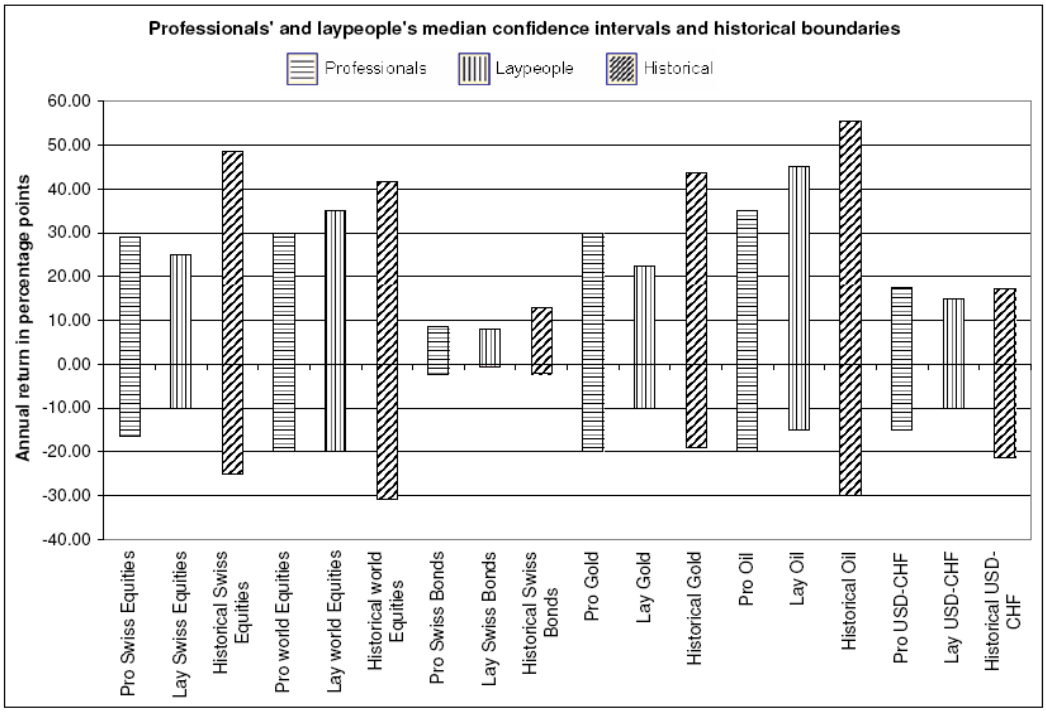
The table 2.1 shows the implied volatilities of the confidence intervals from the professionals and the laypeople respectively for historical returns for each asset class. It also contains the historically implied volatility in each asset class over the last 36 years and the ratio of the participants' implied volatilities and historically implied volatilities. The stars indicate if the differences between implied volatilities from participants' confidence intervals are significantly different than the historically implied intervals in each asset class.

Asset class	Swiss equities	World equities	CHF bonds	Gold	Oil	USD-CHF
Implied volatility professionals	13.08%***	15.38%***	3.23%***	15.85%	18.46%	9.23%
Implied volatility laypeople	12.31%***	15.38%***	2.77%***	10.77%	15.38%	7.69%
Historically implied volatility	22.62%	22.30%	4.62%	19.26%	26.22%	11.84%
Ratio of professionals' implied volatility vs. historically implied volatility	0.58	0.69	0.70	0.82	0.70	0.78
Ratio of laypeople's implied volatility vs. historically implied volatility	0.54	0.69	0.60	0.56	0.59	0.65

* significant at 10% level
** significant at 5% level
*** significant at 1% level

Figure 2-1: Confidence intervals for historical returns

The figure 2.1 shows the median lower and the median upper boundaries for the confidence intervals in the professional and the laypeople sample for historical return estimates of 6 different asset classes as well as the upper and lower boundaries of the realized return of those asset classes over the last 36 years.



Further indication for miscalibration is given by a comparison of the implied volatilities embedded in the confidence intervals and the implied volatility of historical returns. Table 2.1 shows the implied volatilities of the confidence intervals from the professionals and the laypeople respectively for historical returns for each asset class. We see that the implied volatilities in both samples are lower than the historically implied volatilities in all asset classes and the difference is highly significant for equities and bonds. Another evidence for miscalibration is given by the fact that more than 70% of the professionals and 75% of the laypeople provide 90% confidence intervals which are narrower than the historical intervals on each asset class.

An unexpected but interesting result is the fact that the professionals as well as the laypeople have a good feeling for the relative risk of each asset class. The ratios between the historical intervals and the subjective confidence intervals are close to 0.7 and 0.6 respectively in all asset classes as the last two lines in table 2.1 show. Those ratios are calculated by dividing the historically implied volatility by the median implied volatility of the participants' confidence intervals. So both the professional and the laypeople sample are well informed about the relative risk of each asset class or in other words are equally miscalibrated across the historical returns of different asset classes.

2.4.2 Return forecasts

Now we turn to the participants 90% confidence intervals for return forecasts for the year 2006. We asked for return forecasts in Swiss equities, CHF bonds, the participants' own pension plan, and the average Swiss pension plan. The implied volatilities of our professional sample for Swiss equities (3.1%), CHF bonds (1.2%) and the two types of pension plan returns (1.5% for each forecast) are difficult to compare to reasonable benchmarks but we have the impression that those confidence intervals are very narrow relative to the historically implied volatilities in table 2.1. So the participants in our sample express a high conviction about their own forecasting abilities because they choose very narrow upper and lower boundaries in their forecasting intervals.[5] Similar to the estimation of historical returns laypeople provide narrower confidence intervals than professionals but in the forecasting task Mann-Whitney tests confirm that the con-

[5] We acknowledge that the participants provided their answers between May 2006 and August 2006 and it might be the case that they already used the available information for the year 2006. But even then the upper and lower boundaries are on average very close to each other.

fidence interval sizes are significantly different. So the professionals express less confidence into their own forecast abilities than the laypeople and on average they share a more conservative view with respect to downside risk but expect a comparable upside potential.

In line with the observation of Alpert and Raiffa (1982) the participants in both samples express very stable answering patterns when providing 90% confidence intervals for different asset classes. The correlation in the professional sample (laypeople sample) for forecast intervals ranges from 0.72 to 0.92 (0.47 to 0.89) and for the historical intervals the range lies between 0.33 and 0.92 (0.44 and 0.77). Even the correlations between forecasts intervals and historical intervals are always positively correlated with an average of 0.32 (0.18). Providing narrow confidence intervals seems to be a stable trait across individuals regardless of the asset class and the type of estimation.

We can summarize our results so far by saying that the decision-makers of Swiss pension plans as well as a sample of laypeople are miscalibrated when estimating historical returns on financial markets. They also express a high conviction in their forecasting abilities as they provide very narrow confidence intervals for future returns but the effect among professionals is less extreme than across laypeople and professionals have more conservative expectations. Furthermore the participants in both samples are roughly equally miscalibrated across all asset classes and express stable confidence interval patterns when forecasting returns on financial markets. The next section addresses the relationships between individual characteristics and the degree of miscalibration with a linear regression model.

2.4.3 Linear regressions

Across all asset classes in both tasks not only correlations but also Cronbach Alphas are very high. For the professionals (laypeople) forecast intervals they are at 0.88 (0.86) and for the historical intervals at 0.84 (0.84). That allows for an aggregation of the confidence intervals of the different asset classes in each task to define the dependent variables in our linear regression models.[6] The dependent variables are the normalized log values of the participants' 90% confidence intervals for forecasts and historical returns divided into professional and laypeople samples. The first two predictors are mutually

[6] All linear regressions with single asset classes as dependent variables are available on request.

exclusive dummy variables for the different types of education, i.e. holding a degree from university or attendance of practical financial education. Participants with both a university degree and a practical finance education were considered as people with university degree to mitigate double counting. The third predictor reflects a participants' experience in finance or pension plans and can take values from 1 (no experience) to 7 (more than 25 years of experience). The last predictor can also take values from 1 (below 25 years) to 7 (above 65 years) and reflects the range of a participants' age. The last two rows contain the R^2 of the regression models for each type of confidence interval as well as its F-value.

Table 2-2: Linear regression models on confidence interval sizes

Table 2.2 shows the standardized Beta and T values of the predictors in our linear regression models for professionals' and laypeople's forecasts and historical estimations as well as R^2 and F values for the our 4 different linear regression models.

Dependent variables	Professionals' aggregated forecast interval size	Laypeople's aggregated forecast interval size	Professionals' aggregated historical interval size	Laypeople's aggregated historical interval size
Predictors	Standardized Beta Coefficients and T values for 90% confidence interval sizes			
University Degree	0.344 2.600**	0.316 3.427***	0.432 2.641**	0.125 0.741
Practical finance education	0.160 1.210	0.050 0.530	0.235 1.460	-0.057 -0.324
Experience	0.209 1.953*	0.143 1.401	0.360 2.845***	-0.028 -0.142
Age	-0.226 -2.111***	-0.334 -3.360***	-0.465 -3.641***	-0.077 -0.418
R-square	0.139	0.190	0.329	0.031
F	3.234**	5.642***	5.754***	0.285

* significant at 10% level
** significant at 5% level
*** significant at 1% level

Our regression model works well for the professional sample with a R^2 of 13.9% for the forecast intervals and 32.9% for the historical intervals. It also explains 19% of the variation in the confidence intervals for laypeople's forecasts but it fails to relate our predictors to the 90% confidence intervals for historical returns in the laypeople sample. Having a degree from university is significantly related to broader confidence intervals in all asset classes. In contrast practical financial education is not significantly related to the interval sizes. In line with the model from Odean and Gervais (2001) the variable for experience tends to reduce overconfidence as professionals with more financial or pen-

sion plan experience provide significantly broader confidence intervals. Age is related to narrower confidence intervals as older people provide significantly narrower confidence intervals.

To summarize, our regression analysis indicates that older people without a degree from university and with little experience in finance or pension plans provide significantly narrower confidence intervals for returns on financial markets than younger people with a degree from university and more experience.

2.5 Discussion and conclusion

We confirm that people are overconfident in the domain of financial markets but provide new evidence that this is also the case in a sample of decision-makers of Swiss pension plans. So we want to emphasize two practical issues related to overconfidence in the domain of financial markets that might apply to several pension plans and other investors.

First, an overconfidently biased perception of low risks in a volatile asset class like for example equities could increase its perceived attractivity. This might then result in an overweight of that asset class and exposes a pension plan's portfolio more to downside risks but the overconfident decision-makers might be unaware of that risk. De Long, Shleifer, Summers and Waldman (1990) present a theoretical model to demonstrate that noise traders with erroneous stochastic beliefs (like for example miscalibrated investors) take excessive risk and gain less expected utility than rational investors. Having said that we point out that the participants in our sample provided too narrow confidence intervals for all asset classes and not only for more volatile ones so we cannot generalize the argument that high risk assets are overweighted. So further research is needed to address the relationship between overconfidence and the weighting of risky asset in a strategic asset allocation.

The second issue is related to present findings of other authors who demonstrate that overconfidence influences trading decisions of investors. In those studies overconfidence is a drag on performance either because of higher transaction costs due to an increased trading volume (Odean (1999)) or because investors misperceive the true probabilities of market situations and over- or underreact (Daniel, Hirshleifer and Subrahmanyam (1998)). Tactical trading always generates additional transaction costs that

have to be compensated by higher returns but so far academic research rejects the thesis that tactical trading, often referred to as timing, pays off in general. Daniel, Grinblatt, Titman and Wermers (1997) report no systematic timing success for mutual funds and Blake, Lehman and Timmermann (1999) report that UK pension plans have on average no timing skills.

To put our results into perspective Yaniv and Foster (1997) argue that there is a trade-off between accuracy and informativeness when providing confidence intervals. Narrow confidence intervals usually provide more information than large but accurate intervals. We can not rule out that the participants in both of our samples want to provide very informative confidence intervals, especially in the forecasting task, knowing that those might not be totally accurate. We also take in to account that we can not relate the answers of our professionals to the investments of a particular pension plan because we asked all participants to express their personal views in our questionnaire and not the views of their pension plans.

More in depth research is needed to model the relationship between individual characteristics and overconfidence but one straightforward thesis to explain why people tend not to be homogenously miscalibrated is the effect of perceived task difficulty. Task difficulty can not be measured in our questionnaire on an absolute basis. But it might be the case that older people with less education and less experience find it more difficult to come up with appropriate confidence intervals for asset returns on financial markets than younger people with a better education and more financial experience. According to Lichtenstein, Fischhoff and Philips (1982) a higher level of task difficulty is linked to a higher level of miscalibration and this is in line with our observations. However, the unexplained variance in our model indicates that other factors are necessary to explain individual differences in miscalibration in the domain of financial markets, which needs further exploration.

2.6 Literature

Alpert, M., and Raiffa, H. "A Progress Report on the Training of Probability Assessors." in Slovic, P., ed.: Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press, New York, 1982.

Blake, D., Lehman, B., and Timmermann, A. "Asset Allocation Dynamics and Pension Fund Performance." *The Journal of Business,,* 72, (1999), pp.

Brenner, L., Liberman, V., and Tversky, A. "Overconfidence in Probability and Frequency Judgements: A Critical Examination." *Organizational Behaviour and Human Decision Processes,* 65, (1996), pp. 212-219.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. "Measuring Mutual Fund Performance with Characteristic Benchmarks." *Journal of Finance,* 52, (1997), pp.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. "Investor Psychology and Security Market Under- and Overreaction." *Journal of Finance,* 53, (1998), pp. 1839 - 1885.

De Long, B., Shleifer, A., Summers, L., and Waldman, R. "Noise trader risk in financial markets." *Journal of Political Economy,* (1990), pp. 703-738.

Deaves, R., Lueders, E., and Schroeder, M. "The Dynamics of Overconfidence: Evidence from Stock Market Forecasters." *ZEW Discussion Paper No. 05-83*, 2005.

Erev, I., Wallsten, T., and Budescu, D. "Simultaneous Over- and Underconfidence: The Role of Error in Judgement Process." *Psychological Review,* 101, (1994), pp. 519-527.

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. "Probabilistic mental models. A brunswikian theory of confidence." *Psychological Review,* 98, (1991), pp. 506-528.

Glaser, M., and Weber, M. "Overconfidence and Trading Volume." *Sonderforschungsbereich 504*, 03-07, 2003.

Glaser, M., Weber, M., and Langer, T. "On the trend recognition ability of professional traders: Are they better or just more overconfident than lay people?" *CEPR Discussion Paper*, 3904, 2003.

Glaser, M., Weber, M., and Langer, T. "Overconfidence of Professionals and Lay Men: Individual Differences Within and Between Tasks." *Sonderforschungsbereich 504*, 05-25, 2005.

Graham, J., and Campbell, H. "Expectations of Equity Risk Premia, Volatility and Asymmetry." *Working Paper*, 2003.

Graham, J., Campbell, H., and Huang, H. "Investor Competence, Trading Frequency and Home Bias." *NBER Working Paper*, W11426, 2006.

Griffin, D., and Tversky, A. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology,* 24, (1992), pp. 411-435.

Guiso, L., and Jappelli, T. "Information Acquisition, Overconfidence and Portfolio Performance." *CEPR Discussion Paper*, 5901, 2005.

Hilton, D. "The Psychology of Financial Decision-Making: Applications to Trading, Dealing and Investment Analysis." *The Journal of Psychology and Financial Markets,* 2, (2001), pp. 37-53.

Klayman, J., Soll, J., Gonzales-Vallejo, C., and Barlas, S. "Overconfidence: It Depends on How, What and Whom You Ask." *Organizational Behaviour and Human Decision Processes,* 79, (1999), pp. 216-247.

Lichtenstein, S., Fischhoff, B., and Philips, L. "Calibration of Probabilities." in Slovic, D., ed.: Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, 1982.

Locke, P., and Mann, S. "House Money and Overconfidence on the Trading Floor." *AFA 2002 Atlanta Meetings*, 2001.

Menkhoff, L., and Schmidt, U. "The Use of Trading Strategies by Fund Managers: Some First Survey Evidence." *Discussion Paper*, 314, 2005.

Odean, T. "Volume, Volatility, Price and Profit When All Traders Are Above Average." *The Journal of Finance,* 53, (1998), pp. 1887-1934.

Odean, T. "Do Investors trade too Much." *The American economic review,* 89, (1999), pp. 1279-1298.

Odean, T., and Barber, B. "Boys will be boys: Gender, Overconfidence and Common Stock Investment." *The Quarterly Journal of Economics,* 116, (2001), pp. 261-292.

Odean, T., and Gervais, S. "Learning to be overconfident." *The Review of Financial Studies,* 14, (2001), pp. 1-27.

Pearson, E., and Tukey, J. "Approximate means and standard deviations based on distances between percentage points and frequency curves." *Biometrika,* 52, (1965), pp. 533-546.

Russo, E., and Schoemaker, P. "Managing Overconfidence." *Sloan Management Review,* 33, (1992), pp. 7-17.

Soll, J., and Klayman, J. "Overconfidence in interval estimates." *Journal of experimental psychology,* 30, (2004), pp. 299-314.

Yaniv, I., and Foster, D. "Precision and Accuracy in Judgemental Estimation." *Journal of Behavioral Decision Making,* 10, (1997), pp. 21-32.

3. Overconfidence and Active Management: An Empirical Study across Swiss Pension Plans

Abstract

Pension plans in Switzerland favour active management over indexing to implement their strategic asset allocation. Empirical surveys show, however, that their success in the past has been below expectations as the median performance of Swiss pension plans in domestic and international equities is below market indices even gross of fees. The results of this paper's survey across decision-makers of Swiss pension plans sheds some light on why active management is still so popular across Swiss pension plans. On average the participants in the sample are prone to the better-than-average-effect. A majority expects their managers and their overall pension plan to outperform the other survey participants in the future. The subjective perceptions of the own skill level relative to the competitors can explain the popularity of active management across Swiss pension plans.

3.1 Introduction

The goal to be above average is deeply rooted in human nature and can be observed in many different domains. Studies in psychology show that people also tend to have the illusion that they are capable of delivering above average performances in various tasks (Taylor and Brown (1988)) despite having no adequate means to compare themselves with a representative average. This is often referred to as one facet of overconfidence called the "better-than-average-effect" and financial markets are no exception with respect to such an effect across investors.

In contrast to other domains an adequate measure of the average performance is easily available on financial markets as data about various broad market-capitalized indices are freely accessible for investors. However beating the average is difficult. As Sharpe (1991) demonstrated, active management is a zero-sum game and empirical evidence shows that only a minor percentage of all investors are able to beat the performance of broad market indices net of fees in the long run. On average Swiss pension plans are no exception as they have failed to beat equity market index returns, too. Nevertheless the lure of potential outperformance ensures active management remains very popular across Swiss pension plans and the decision-makers express surprisingly high confidence to beat their peers in the future. The better-than-average-effect can explain the gap between Swiss pension plans' expectations about their future success and their realized historical success.

The contribution to research of this paper is twofold. First, evidence is presented that the implementation of the strategic asset allocation with active managers is more popular but historically not more successful than indexing from a risk-return perspective across Swiss pension plans. In the last couple of years Swiss pension plans achieved on average lower returns than broad market indices in domestic as well as international equities. The second contribution deals with expectations concerning the future performance of their own pension plans in a sample of decision-makers at Swiss pension plans. Their answers to the author's questionnaire reveal that they expect their own pension plans as well as their managers to outperform the other participants' in the sample and to be able to select above average managers. Such optimistic expectations are puzzling when the average historical performance is taken into account and indicates that the participants in the sample are indeed prone to the better-than-average-effect.

It is far beyond the scope of this paper to take a general view on the advantages and disadvantages of active management for Swiss pension plans. Successful active management depends on an array of circumstances like goals, skills, expectations, risk-aversions and constraints. So there is no way to argue for an optimal degree of active management in general because there is too much heterogeneity across Swiss pension plans with respect to those different circumstances. However this paper argues that susceptibility to the better-than-average-effect can bias the judgment of the own skills relative to other investors and therefore can explain the heavy reliance of the Swiss pension plan industry on active management in equities despite the lack of success in the past.

The remainder of the paper is organized as follows. Section II outlines the environment and framework Swiss pension plans are operating in. Section III reviews related research on active management with a special focus on pension plan performances and on the better-than-average-effect. Section IV describes the data including the answers from the author's questionnaire for the decision-makers of Swiss pension plans. Section V presents the performance of Swiss pension plans relative to market indices and the proneness to the better-than-average-effect in the sample. Together with the conclusions, section VI discusses implications of the results and addresses alternative explanations for the popularity of active management across Swiss pension plans.

3.2 Swiss pension funds

A sample of Swiss pension plans is distinct from previous studies about private investors' or mutual funds' preferences and performances of active management because the participants in this sample not only bear responsibility for their own investments but for the retirement savings of the employees in Switzerland. According to a study of the Swiss National Bank for the year 2004 (see SNB (2006)) roughly half of all the peoples' wealth in Switzerland, around CHF 500 billion, is managed in the second pillar, i.e. in the hands of Swiss pension plans.[7] So it is in the interest of the working people in Switzerland that Swiss pension plans take wise and prudent financial decisions.

The pension funds under examination are segregated funded occupational pension schemes from the private and the public sector in Switzerland. Throughout their work-

[7]. In September 2006 estimates for the wealth in Switzerland's second pillar are at CHF 650 Bn Swisscanto, 2006, Schweizer Pensionskassen 2006, (Swisscanto)..

ing life Swiss employees of companies or public institutions provide contributions to their pension funds which accumulate over time and then are used to fund retirement payments to pensioners. It is the goal of the pension funds to manage those inflows in a way to guarantee the payments to the pensioners without imposing high contributions to the current employees. The pension funds are provided by a public or private employer (the sponsor) and are either defined benefit or defined contribution plans or a hybrid of those two. Defined benefit plans offer the employee guaranteed payments in retirement from the sponsor usually defined as a percentage of the employee's final salary. The risk of funding such a guaranteed payment is borne by the sponsor. By contrast defined contribution plans convert the value of an employee's savings in the pension fund into an annuity at retirement. In a defined contribution plan the employee bears the risk. In Switzerland there are many hybrid schemes that contain elements of both basic types of schemes. Whether it is the employee or the sponsor that carries the higher burden of risk, there is no historical evidence to suggest that either type of scheme has more or less success using active management, nor that there are significant differences in the level of popularity of active management between the two (Frauenlob (1998)).

In Switzerland the law forces the board of trustees to take decisions about the strategic asset allocation of their pension plan and they cannot delegate this responsibility. However they can rely on third party knowledge in the elaboration of the strategic asset allocation. Members of the board of trustees are not necessarily investment professionals but rather people who represent stakeholders of the sponsor and have been elected into the board to lobby their voters' interest. The chosen asset allocation of every pension plan has to be in line with the Swiss regulatory framework.[8] The implementation of the chosen asset allocation in a particular pension plan is usually delegated to an investment committee which consists of delegates from the board of trustees and / or investment professionals. Most of the investment committees of Swiss pension plans consist of three to six members and meet between four and ten times per year. The main responsibility of such an investment committee is to appoint asset managers and to take tactical decisions within the guidelines defined by the board of trustees. Investment committees typically decide whether the day-to-day management is delegated to an external asset

[8] Verordnung über die berufliche Alters-, Hinterlassenen- und Invalidenvorsorge (BVV 2), articles 50 to 59.

manager or whether in-house managers are hired. They also have the option to delegate certain tactical decisions to in-house managers. So the decision about the role of active management in the implementation of the strategic asset allocation lies in the hands of either the investment committee members or the in-house managers.

In order to analyze explanations for the popularity of active management across Swiss pension plans a sample of investment committee members and internal managers seems to be appropriate. However, the important role played by consultants and the advice they give to pension fund professionals cannot be overlooked. Indeed, it would be inaccurate to suggest that all decisions are taken independently by the investment committees. According to a survey from Lusenti (2003) across 195 public and private Swiss pension plans 38% use third party advice in strategic decisions. Only 20% of the plans in the same sample indicate that they never work with external consultants. Nevertheless an investment committee ultimately bears the responsibility so it is their decision that matters.

3.3 Related research

This paper does not go into detail on the ongoing debate about the success achievable through active management but a few introductory phrases shed some light on the basic framework in which the results in this paper have to be viewed. Then the historical performance of pension plans in various countries relative to market indices will be addressed. The last part of this section focuses on the better-than-average-effect whose roots are in the psychological literature.

Decades ago the discussion started with the development of the CAPM (Sharpe (1964)), the Efficient Market Hypothesis by Fama (1970) and its weaknesses pointed out by Roll (1977) or Grossman and Stiglitz (1980). Since then the basic question has been whether a broad market index represents a better investment opportunity than an actively-managed portfolio. Carhart (1997), Daniel, Grinblatt, Titman and Wermers (1997) and Malkiel (2004) all present empirical evidence that most mutual fund managers are unable to deliver a persistent outperformance versus a broad market index net of fees. By the same token there is no doubt that some active managers delivered an outperformance versus market indices (Siegel, Kroner and Clifford (2001)) – maybe due to luck, maybe due to skill - and investing in such managers would have yielded a significantly

higher return than purely indexed exposure to markets. So how did the pension plan industry perform so far?

Frauenlob (1998) describes investment strategies for Swiss pension plans and reports that pension plans with a more active implementation of the strategic asset allocation do not perform better than other pension plans. But he points out that most of the pension plans did not outperform broad market indices net of fees in different asset classes and over different time periods. Characteristics of a pension plan such as size, assets under management or pension scheme type offered no explanation for the degree of active management in his study. The results of previous studies in other countries about the success of active managers for pension funds are mixed but the majority does not favor active management. Lakonishok, Shleifer and Vishny (1992) report that from 1983 to 1989 most of the active equity managers in a sample of 769 funds that managed assets of US pension plans delivered an underperformance and in total they underperformed the S&P 500 by 1.3% p.a. gross of fees. In contrast Berzins and Trzcinka (2005) report that a sample of 549 US equity portfolios at US pension plans outperformed the S&P 500 by a mean of 0.65% p.a. gross of fees from 1993 to 2003. For a sample of UK pension funds Blake, Lehman and Timmermann (1999) conclude that the pension plans have no selection or timing skills. In a later report Blake, Lehman and Timmermann (2002) demonstrated that in a sample of 306 UK pension funds over the period from 1986 to 1994 the median pension plan underperformed a broad market index by 0.15% p.a. in UK equities and 0.06% p.a. in international equities gross of fees. On a total portfolio level 138 UK pension plans were able to realize an outperformance versus an aggregated benchmark and 168 delivered an underperformance. The authors further observed a low cross-sectional variability in returns indicating that the UK pension plan managers tend to take low relative risk versus market indices. In a sample of 2175 equity portfolios of UK pension funds Tonks (2001) reports an average underperformance of 0.006% p.a. against the FTSE All Share Index gross of fees from March 1983 to December 1997. However he points out that there is significant persistence across outperformers as well as across underperformers. Stanko (2003) reports that on average active management added value to the performance of 21 public pension plans in Poland in the period from June 1999 to March 2003 on an overall portfolio level but no details about single asset classes are available.

From the realized performance of pension plans around the world we now turn to the psychological literature which offers various insights into the better-than-average-effect. The better-than-average-effect is one aspect in the field of overconfidence which is a complex phenomenon with various facets. In the domain of financial markets Glaser and Weber (2007) differentiate between 4 different manifestations of overconfidence; miscalibration, better-than-average-effect, illusion of control and overoptimism. This paper only concentrates on the better-than-average-effect because it is directly observable in a questionnaire, it is relevant in the domain of financial markets as Graham, Campbell and Huang (2006) outline and it serves as one explanation for the popularity of active management across Swiss pension plans.

The better-than-average-effect was first observed in psychology. Svenson (1981) notes that more than 80% of the participants in a survey believe themselves to be above average with respect to their driving skills. Taylor and Brown (1988) demonstrate that people generally tend to be prone to the better-than-average-effect in many fields as they believe their skills and abilities to be above average despite the absence of reliable information about their true level of skill compared to others. In a later paper (Taylor and Brown (1994)) they confirm their thesis and point out that being overly optimistic is usually increasing the well-being of an individual. So from a psychological standpoint being prone to the better-than-average can be seen as rather healthy and definitely non-pathological but in the domain of financial markets it can lead to overoptimistically biased expectations about the own abilities and the own performance in the future. Camerer and Lovallo (1999) present evidence from an experimental market that roughly 70% of the participants in their sample are prone to the better-than-average-effect in a market entry game when relying on the subjective perception of the own skills relative to the competitors in the game. They further report that the participants on average lose money due to their propensity to the better-than-average-effect and their overestimation of their own skills and so this results in financial decisions that lower their wealth.

In a theoretical model De Long, Shleifer, Summers and Waldman (1991) show that overconfident investors can survive on financial markets but tend to take more risk and gain less expected utility than rational investors. Kyle and Wang (1997) show with their model that employing overconfident managers can be the best strategy, in the context of game theory, as both participants face a prisoner's dilemma. However the choice of an

overconfident manager does not lead to an efficient outcome for both players. It is acknowledged that these models are applied directly to investors and not to decision-makers who have the choice of delegating the portfolio management to an external asset manager. However the point is to show that overconfident market participants can theoretically survive on financial markets but do not achieve the best risk-adjusted performance. Lakonishok, Shleifer and Vishny (1992) argue that overconfidence could lead to a preference for active management in a pension plan despite the fact that it might deliver returns below a market index. They point out that overconfidence about selection skills used to identify good active managers can explain the high percentage of active management within US pension plans.

Empirically Odean and Barber (2000) and Guiso and Jappelli (2005) demonstrate in samples of private investors that high trading volumes can reduce the performance results substantially due to increased trading costs. It seems fair to assume that the trading costs for Swiss pension plans are lower than for private investors but basically higher trading volumes cause more transactions costs for pension plans as well.

There is no doubt that it is very difficult to beat the average performance on financial markets in the long run. But maybe some investors are not really aware of it because the better-than-average-effect biases their perceptions about the own skill level and the thoughness of competition. In the next sections it will be analyzed if this is also the case for Swiss pension plans.

3.4 Data and methods

In this paper the focus is on the implementation of a strategic asset allocation with active managers instead of indexers across Swiss pension plans. The availability of index instruments today gives the decision-makers of Swiss pension plans a true choice between active managers and indexers when implementing the strategic asset allocation for traditional asset classes like equities or bonds. The choice of a strategic asset allocation itself can in fact also be regarded as active management as the board of trustees of a pension plan has to take active decisions. However for the strategic asset allocation of a Swiss pension plan there is no market index that could be used as a default portfolio so there is no true choice between active managers and indexers. Therefore the focus is only on the performance of Swiss pension plans within asset classes and not across total

pension plan portfolios. Timing skills of pension plans will also be excluded from the analysis in this paper because meaningful data is not available.

In order to have enough observations and a true choice between active management and indexing the paper only analyzes the performances in three asset classes. Those are domestic and international equities as well as CHF bonds. We then compare the pension plans' average and median performances in those asset classes with well established market indices. For domestic equities the Swiss Performance Index (SPI) is applied as a benchmark, for international equities the MSCI World Index family is used, and for CHF bonds the Swiss Bond Index (SBI). Because Swiss pension plans are tax exempt, the loss due to taxes on dividends is in most of the cases marginal and so a total return index is a fair comparison. It might be the case that many pension plans in the sample do not use those indices as their benchmarks. But the choice of a benchmark is also an active decision that investment committees or managers have to take and the MSCI index family, the SPI and the SBI can be seen as reasonable and investable default solutions for those asset classes.

To collect data about the absolute and relative performance of Swiss pension plans within asset classes the paper relies on two different types of empirical surveys. First, on annual studies from Lusenti Partners (Lusenti (2007), Lusenti (2006) and Lusenti (2005)) which are based on a sample of 130, 123 and 123 Swiss pension plans respectively. These surveys present mean and median realized annual performances across different asset classes which can be compared to market index returns. Second, on the performance data from the ASIP Performance Comparison across 60 to 73 Swiss pension plans over different annual periods (ASIP (2000-2006)). The availability of data about returns of pension plan portfolios[9] in different asset classes in the ASIP survey allows for a comparison between each pension plan's performances within the same asset class relative to a market index. While the Lusenti surveys directly collect all their information from the pension plans, ASIP uses financial data from the global custodians of the participating pension plans. This guarantees for generally-accepted calculation methods in the ASIP data and allows for a more homogenous set of performance data to

[9] A pension plan portfolio is defined as a single mandate to an in-house or external asset manager within a certain asset class or a consolidation of single mandates of a pension plan within an asset class. There is no further data available about the performances of each single mandate that is included in the consolidation of mandates of a pension plan.

compare. ASIP relies on returns gross of fees whereas Lusenti requests data net of fees. This has to be taken into account when interpreting the results.

This paper does not assume that one of those surveys is singularly representative for Swiss pension plans because both include only a small part all pension plans in Switzerland.[10] However in terms of assets under management the samples from Lusenti include roughly CHF 200 billion in each year while the samples in ASIP contain assets of roughly CHF 80 billion in 2000 to CHF 160 billion in 2006. This reflects one third and one quarter respectively of total asset under management in the second pillar of Switzerland. The ratio between participating pension plans and assets under management shows that rather large pension plans participate in the ASIP and the Lusenti survey so a size bias cannot be ruled out. But the results of this paper are on the conservative side as larger pension plans tend to index more assets than smaller funds according to Lusenti (2003). So this paper tends to underestimates the popularity of active management across Swiss pension plans. There is a self selection bias in the samples because there is no obligation to participate in either the Lusenti or the ASIP surveys. It is plausible that only successful and above average pension plans are incentivised to participate in those surveys. The incentive to participate in the ASIP survey might be additionally limited by the fact that a fee is charged for the participation. In light of this bias the results with respect to historical success with active management rather overestimate the success of the overall pension plan industry in Switzerland. There is a certain overlap in the samples because some pension plans participate in the Lusenti as well as in the ASIP survey but is not quantifiable because data about the participants in the Lusenti study is not public. An analysis between the performance of defined contribution and defined benefit plans is not valid in this paper due to the low number of defined contribution plans in the sample but in line with Frauenlob (1998) no indications for differences appeared in the available data.

In order to analyze the better-than-average-effect across decision-makers of Swiss pension plans the author constructed a specific questionnaire that has been distributed to large sample of investment committee members and managers at Swiss pension plans. In total 584 questionnaires have been distributed and 132 have been returned. This cor-

[10] In 2004 BFS, Bundesamt für Statistik, 2004, Pensionskassenstatistik, (Bundesamt für Statistik BFS, Neuchatel). counted 2935 pension plans in Switzerland. There is a decreasing trend in the last 10 years.

responds to a response rate of 22.6%. A majority of the participants' pension plans are represented in either the Lusenti or the ASIP survey. 24 questionnaires contained incomplete information and therefore have been excluded from the analysis. As a result the professional sample consists of 108 participants (56 investment committee members, 52 managers) who represent a total of 39 Swiss pension plans (22 Defined Benefit, 8 Defined Contribution, 2 Hybrids, and 7 unspecified).

To measure the better-than-average-effect the participants were given two different tasks in a questionnaire. In the first task they were asked about their expectations with respect to their own pension plans' future success relative to the other participants in the sample. The questions in that part were about i) the chances of their pension plan to find above average active managers in the future, ii) the likelihood of their pension plan to achieve an above average risk-adjusted return in the future, and iii) the chances their internal and external managers would outperform the other active managers of Swiss pension plans in the sample in the future. An exact wording of the questions is provided in appendix I. To answer those questions the participants had to tick a box on a Likert scale from 1 (clearly below average) to 7 (clearly above average). The possibility of 4 (average) is included as many participants probably would choose this option. The participants were well informed about their competitors in this sample as this was explicitly stated in the questionnaire. In the second part the participants had to provide subjective confidence intervals to forecast the returns of different asset classes for the calendar year 2006 that contain the realized return with a probability of 90%. A comparison across the lower and upper boundaries of those confidence intervals provides information about the participants' expected returns of pension plan portfolios.[11] In addition they had to provide a self evaluation of the accuracy of their own forecasts versus those of the other participants.

[11] No comparisons of implied returns from subjective confidence intervals are made because the mean of a confidence interval could be a biased result because the participants might have a skewed distribution in mind.

3.5 Results

3.5.1 Historical performance of Swiss pension plans

In the annual surveys from Lusenti for the years 2004 to 2006 and ASIP from 2000 to 2006 there is no indication of outperformance versus market indices across Swiss pension plans. Table 3.1 provides an overview of the median performances of Swiss pension plans in different asset classes according to those two surveys. The median was favored over the mean to mitigate outlier effects.

Table 3-1:Past performance of Swiss pension plans
For each asset class the top rows show the sample median returns of the Lusenti Partners survey and the second rows the sample median returns for the ASIP Performance Comparison. The last rows each present the returns of market indices for the asset classes. The last 2 rows in the table present the returns of the MSCI World Index in CHF inclusive Switzerland and exclusive Switzerland. The reason to show both indices is that the treatment of Swiss equities in a world equity index is different across Swiss pension plans and also across the surveys from Lusenti Partners and ASIP. Some plans include Swiss equities within the universe of world equities and others do not.

	Median annual returns from Lusenti and ASIP surveys and market indices						
Year	2000	2001	2002	2003	2004	2005	2006
Sample size for Lusenti / ASIP	na / 50	na / 63	na / 65	na / 70	123* / 71	123** / 73	130*** / 72
AuM for Lusenti / ASIP in bn CHF	na / 80	na / 80	na / 75	na / 100	215 / 145	190 / 145	212 / 161
CHF bonds							
Lusenti median	na	na	na	na	3.4%	2.7%	0.3%
ASIP median	3.7%	3.9%	9.9%	1.6%	3.6%	2.7%	0.0%
Swiss Bond Index (SBI) TR domestic"	3.4%	4.4%	10.8%	1.5%	4.6%	3.6%	-0.3%
Swiss equities							
Lusenti median	na	na	na	na	5.9%	34.0%	20.0%
ASIP median	11.4%	-21.5%	-26.9%	21.7%	6.3%	35.2%	20.1%
Swiss Performance Index (SPI) TR	11.9%	-22.0%	-25.9%	22.1%	6.9%	35.6%	20.7%
World equities							
Lusenti median	na	na	na	na	5.5%	27.0%	13.0%
ASIP median	-11.1%	-16.3%	-33.6%	19.1%	5.4%	27.0%	11.2%
MSCI World Index in CHF TR	-11.9%	-14.5%	-33.0%	19.6%	6.0%	27.5%	11.8%
MSCI World Index ex Switzerland in CHF TR	-12.4%	-14.3%	-33.3%	19.6%	6.0%	27.3%	11.5%

" The SBI domestic index covers only CHF bonds from Swiss companies whereas the SBI general reflects the whole universe of CHF denominated bonds. Most of the Swiss pension plans use the SBI domestic as a benchmark.
* Only 123 out of 174 Swiss pension plans in the sample provided an answer to this question
** Only 123 out of 162 Swiss pension plans in the sample provided an answer to this question
*** Only 130 out of 172 Swiss pension plans in the sample provided an answer to this question

The first three rows show the median annual returns according to the Lusenti and the ASIP survey as well as the return of the Swiss Bond Index (SBI) domestic. Rows four to six and seven to ten contain the median annual returns for Swiss and international equities of the pension plans and for the market indices. Except for the median returns in the ASIP sample for CHF bonds in 2000 and 2003, Swiss equities in 2001 and foreign equities in 2000 and 2006, Swiss pension plans have never outperformed the corre-

sponding market index neither net of fees (Lusenti) nor gross of fees (ASIP). In most of the years the median from ASIP underperforms the annual return of the market index by roughly 0.5% to 1.5%.

An analysis of all portfolios from Swiss pension plans that participate in the ASIP survey confirms that most portfolios in the asset classes Swiss equities, world equities and CHF bonds do not outperform a market index gross of fees. Table 3.2 shows in columns one, three and five the rank of a market index with respect to the return for each asset class within the universe of portfolios across all pension plans in the ASIP survey over different time periods. For example the SPI's return over the last 36 months was ranked 14 within a universe of 46 portfolios in the asset class Swiss equities. So 32 of all the portfolios delivered a lower return than the SPI gross of fees and 13 had a higher return. The returns of indices in all analyzed asset classes are always in the better half of the ASIP pension plan universe except for CHF bonds in the period from June 05 to June 06.

One might argue that the pension plans in the samples from ASIP manage the asset classes with a lower volatility than the market indices and therefore achieve better risk-adjusted returns. The columns two, four and six of table 3.2 list the rank of the market indices' Sharpe Ratios for each asset class for the last 36 and 60 months.

Table 3-2: Rank of market indices in the pension plan universe

The table 3.2 lists the ranks of market indices with respect to returns and Sharpe Ratios for three different asset classes, Swiss and world equities and CHF bonds in the sample of the ASIP Performance Comparison between different time periods. Each row shows the rank of an index for a past time period and the number of pension plan portfolios in the sample. To compute the rank total return indices are used. For Swiss equities the Swiss Performance Index (SPI), for world equities the MSCI World Index and for CHF bonds the Swiss Bond Index (SBI) domestic. The risk free rate is the 12 month money market return according to the Swiss National Bank (SNB) for the period from June 30th, 2003 (June 2001) to June 30th, 2006 and is 0.78% (1.47%) annualised.

Time period	Return Rank of Swiss Performance Index (SPI)	Sharpe Ratio Rank of Swiss Perfor-mance Index (SPI)	Return Rank of MSCI World Index	Sharpe Ratio Rank of MSCI World Index	Return Rank of Swiss Bond Index (SBI) domestic	Sharpe Ratio Rank of Swiss Bond Index (SBI) domestic
June 2005 to June 2006	12 / 50	na	24 / 48	na	37 / 49	na
June 2004 to June 2006	10 / 50	na	23 / 46	na	7 / 48	na
June 2003 to June 2006	14 / 46	19 / 41	18 / 44	16 / 41	10 / 46	17 / 44
June 2001 to June 2006	15 / 39	9 / 28	13 / 35	11 / 32	5 / 39	7 / 39

Those ranks are always in the better half of the overall universe in the sample so it is not the case that pension plans delivered better risk-adjusted returns compared to market indices in these asset classes.[12]

Table 3.1 and 3.2 do not differentiate between active portfolios and indexed portfolios so it might be the case that many indexed portfolios slightly underperform the SPI and therefore median return is below the index return. Nonetheless this argument is not valid because of two different reasons. First, a report by Lusenti (2003) indicates that most of the Swiss pension plan assets in various different asset classes are managed actively. Out of 110 plans who answered that question 65 plans report to index between 0% and 10% and further 18 plans index between 10% and 25% of their total assets. On the other hand only 14 plans report to index at least 50% of their total assets. Lusenti (2003) further indicates that the percentage of indexed assets increases with the size of pension plans. As this paper deals with samples of large pension plans it rather underestimates the popularity of active management across Swiss pension plans. For a comparison Ennis (1997) reports that around 35% of the assets of US pension plans are indexed and that the trend is increasing. Lusenti (2003) further reports that there is no significant difference between Swiss and international investments with respect to the popularity of active management across Swiss pension plans.

Table 3.3 contains the second reason and shows that active management in equities - but not in CHF bonds - is dominant in the sample of pension plans whose decision-makers also completed the author's questionnaire about the better-than-average-effect. As can be seen in table 3.3 and graphically in figure 3.1 the majority of all equity portfolios are managed actively in each year. A portfolio is defined as actively-managed if its tracking error was above 0.3% per annum for Swiss equities and CHF bonds and above 0.5% per annum for international equities.[13] With this simple criterion one cannot differentiate between true indexers and active managers who did not spend their active risk budget. But those tracking error levels seem to be a simple and conservative approximation. As benchmarks to calculate the relative Alpha and the tracking error we used market indi-

[12] The risk free rate is the 12 month money market return according to the Swiss National Bank (SNB) for the period from June 30th, 2003 (June 2001) to June 30th, 2006 and is 0.78% (1.47%) annualised.

[13] For Swiss equities it is easier to replicate the market index (SPI) than for world equities because the universe of shares is much smaller. That is why different levels of tracking errors for Swiss and world equities are used to define active portfolios. The relatively homogenous universe of CHF bonds in the SBI domestic explains the low tracking error limit to define active portfolios.

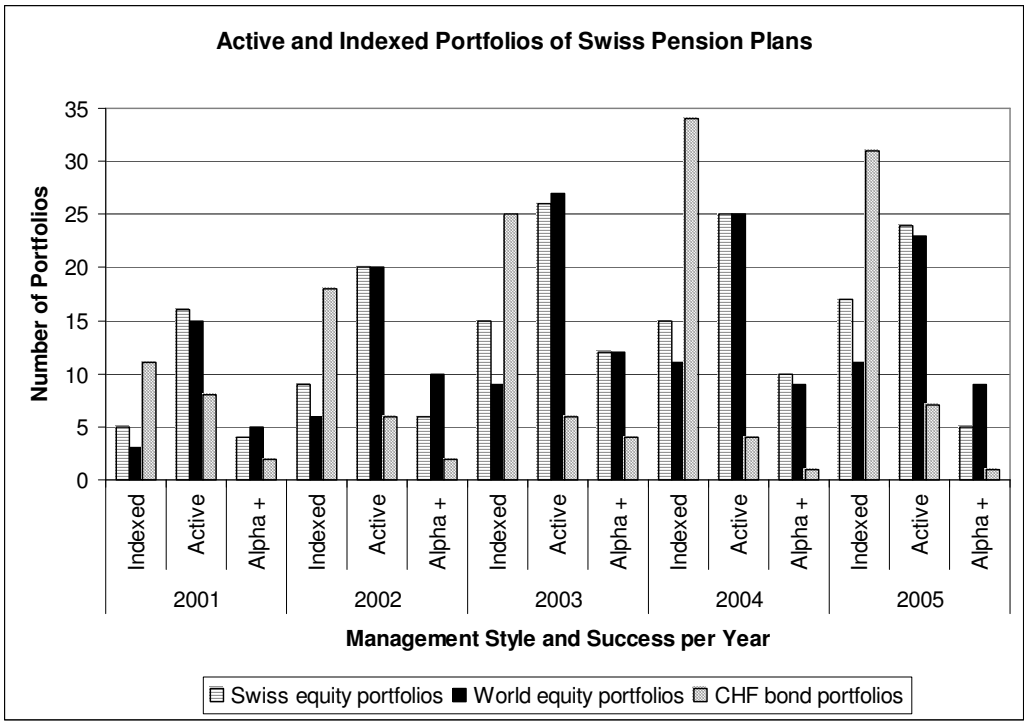
ces. For Swiss equities the Swiss Performance Index (SPI), for World equities the MSCI World Index in CHF and for CHF bonds the Swiss Bond Index (SBI). We note that the percentage of indexed portfolios has steadily grown between 2001 and 2005 but active management in equities is still much more popular. For CHF bonds the picture looks different. This can be seen in the last row of table 3.3, which demonstrates that the pension plans in this sample apply passive management, or at least a type of management that resulted in a very low tracking error to the SBI.

Table 3-3: Popularity and performance of active management
The table 3.3 contains the number of indexed and active portfolios of all the pension plans that participated in the author's questionnaire for each year from 2000 to 2005. Active portfolios are defined as portfolios with a tracking error above 0.3 for Swiss equities and CHF bonds and with a tracking error above 0.5 for world equities. The success rate of the active mandates called "Alpha" in the right columns for every year shows how many of all active portfolios performed better than a corresponding market index. As benchmarks to calculate the relative Alpha and the tracking error we used market indices. For Swiss equities the Swiss Performance Index (SPI), for World equities the MSCI World Index in CHF and for CHF bonds the Swiss Bond Index (SBI). The first row shows the numbers for equities Switzerland and the next two rows contain the same information for world equities and CHF bonds.

Year	2001			2002			2003			2004			2005		
Asset class	Indexed	Active	Alpha	Indexed	Active	Alpha	Indexed	Active	Alpha	Indexed	Active	Alpha	Indexed	Active	Alpha
Swiss equity portfolios	24%	76%	25%	31%	69%	30%	37%	63%	46%	38%	63%	40%	41%	59%	21%
World equity portfolios	17%	83%	33%	23%	77%	50%	25%	75%	44%	31%	69%	36%	32%	68%	39%
CHF bond portfolios	58%	42%	25%	75%	25%	33%	81%	19%	67%	89%	11%	25%	82%	18%	14%

Table 3.3 supports the other results as it shows that most of the actively-managed equity portfolios from pension plans that completed the better-than-average-questionnaire have not beaten the annual performance of an equity market index in the years 2001 to 2005. The columns labeled "Alpha" contain the number of portfolios with a return above the market index in a given year. For equities the success rate, i.e. the number of portfolios that have higher annual returns than the market index, varies but is never above 50% except. Those results indicate that the Swiss pension plans of the decision-makers who completed the better-than-average-questionnaire cannot outperform market indices on average and therefore achieve similar performances to the Lusenti and the ASIP surveys.

Figure 3-1: Active and indexed portfolios of Swiss pension plans
Figure 3.1 shows the same results like table 3.3 graphically.



We can summarize this section by saying that the implementation of the strategic asset allocation with active management is much more popular than indexing across Swiss pension plans. But the evidence about the median performance of Swiss pension plans with domestic and international equities does not support the dominance of active management. The median performance across two well established Swiss pension plan surveys is below market indices' returns and more often than not active portfolios fail to beat the performance of a broad market index even gross of fees.

3.5.2 Better-than-average-effect

This paper argues that a biased view on the own skill level relative to other investors could lead to a heavy reliance on active management. It is now analyzed if decision-makers of Swiss pension plans are really prone to such a better-than-average-effect.
To report the susceptibility to the better-than-average-effect we must first have a look at the participants' expectations concerning their current manager's chances for future success, their own manager selection skills and their pension plan's chances to outper-

form the other participants in the sample. In one part of the questionnaire the participants had to tick a corresponding box on a Likert scale between 1 (clearly below average) to 7 (clearly above average) including 4 (average) to express their opinion. Each column in table 3.4 contains the participants' mean, the t value, the median and the standard deviation (rows three to six) for each one of those questions. The sample mean is above 4 (average) for all questions and one sample t tests reveal significant differences between the sample means and the answering option "average" at the 1% level for all questions except the outlook to find above average active managers in the future which is only significant at the 10% level. This is evidence that the sample is on average prone to the better-than-average-effect in the domain of judging their own and their managers' abilities on financial markets.

Table 3-4: Perceived chances for active management
Each column lists the summary statistics of the participants' answers from the questionnaire. The participants had to tick a box with values from 1 (clearly below average) to 7 (clearly above average) to indicate how they perceive the chances of their own pension plan compared to the other participants' pension plans in the sample. A value of 4 indicates average. Rows 8 to 14 list the percentage of participants who ticked the different answering boxes. Missing values occur either because a participant did not answer the question or because a participants' pension plan does not work with internal or external managers. The latter case occurs much more often.

	Own pension plan's chances to find above average active managers	Current internal active managers' chances to outperform the other internal managers the sample	Current external active managers' chances to outperform the other external managers the sample	Own pension plan's chances to outperform the other pension plans in the sample
Valid	104	74	91	105
Missing	4	34	17	3
Mean	4.17*	4.55***	4.37***	4.57***
T value	1.69	5.74	5.26	6.53
Median	4	5	4	5
Std. Deviation	1.05	0.83	0.68	0.90
Percentages of answers				
% clearly below average	1.92	0.00	0.00	0.95
% below average	3.85	1.35	1.35	0.00
% slightly below average	14.42	5.41	5.41	4.76
% average	43.27	39.19	39.19	43.81
% slightly above average	28.85	47.30	47.30	39.05
% above average	6.73	4.05	4.05	8.57
% clearly above average	0.96	2.70	2.70	2.86

* 10% significance level
** 5% significance level
*** 1% significance level

Rows eight to fourteen in table 3.4 contain the percentage of chosen answers by the participants. The answering options "average" and "slightly above average" are chosen

more often than all three options below average and this is again evidence that the decision-makers in our sample are on average susceptible to the better-than-average-effect. More than 93% of the participants believe to have at least average internal and external managers and to perform at least on an average level compared to the other pension plans in the sample. The first two rows show the sample size and the number of missing participants. A participant's answer is missing either because she refused to answer or because he does not apply internal or external management. The latter case applies in most cases.

Correlations between the answers of the decision-makers of Swiss pension plans across those four questions provide information about the generality of the better-than-average-effect across the participants. Spearman rank correlations are applied because the assumption that the participants' answers are close to normally distributed cannot be made. Higher correlations indicate that a participant expresses optimistic views across all four questions. The Spearman rank correlations indicate a high level of general proneness to the better-than-average-effect because all the correlations are significant at the 1% level and range from 0.306 to 0.541.

In a second step we analyze if the participants in this sample are also prone to the better-than-average-effect when forecasting returns of different asset classes themselves. The six columns in table 3.5 each represent an asset class for which the participants had to judge the accuracy of their own return forecast relative to the other participants in the survey. The same Likert scale like in table 3.4 was applied. The means for all asset classes in table 3.5 are above 4 but one sample t tests provide evidence that the differences are not significantly different from 4 (row four). This is evidence that the participants are not significantly prone to the better-than-average-effect in the task of forecasting returns themselves. The only exceptions are equities and the own pension plan returns as the participants judge their own forecasts to be significantly above average. To be able to forecast the returns of the own pension plan above average looks reasonable as the participants probably have better knowledge about the potential return of the own pension plan. Unsurprisingly the option "average" was by far the most chosen one in the questionnaire. The option "slightly above average" is more popular than the option "rather below average" but the differences are much smaller in the forecasting task than in the tasks about a pension plans chances for future success.

Table 3-5: Better-than-average-effect in the task of forecasting returns
Each column lists the summary statistics of the participants' answers from the questionnaire. The participants had to tick a box with values from 1 (clearly below average) to 7 (clearly above average) to indicate how they judge the accuracy of their own forecasts in each asset class compared to the other participants in the sample. A value of 4 indicates average. Rows 8 to 14 list the percentage of participants who ticked the different answering boxes. Missing values occur because a participant did not answer the question.

Asset Class	Swiss Equities in General	Pension Plan Swiss Equities	CHF Bonds in General	Pension Plan CHF Bonds	Own Pension Plan	Average Pension Plan
Valid	90	86	88	84	90	86
Missing	18	22	20	24	18	22
Mean	4.12*	4.10	4.08	4.10	4.26***	4.07
T value	1.69	1.38	1.12	1.18	3.20	0.90
Median	4.00	4.00	4.00	4.00	4.00	4.00
Standard deviation	0.68	0.70	0.66	0.74	0.76	0.72
Percentages of answers						
% clearly below average	0.00	0.00	0.00	0.00	0.00	0.00
% below average	0.00	1.16	1.14	1.19	0.00	1.16
% rather below average	13.33	12.79	12.50	14.29	11.11	13.95
% average	65.56	63.95	65.91	63.10	60.00	66.28
% rather above average	16.67	18.60	18.18	16.67	21.11	13.95
% above average	4.44	3.49	2.27	4.76	7.78	4.65
% clearly above average	0.00	0.00	0.00	0.00	0.00	0.00

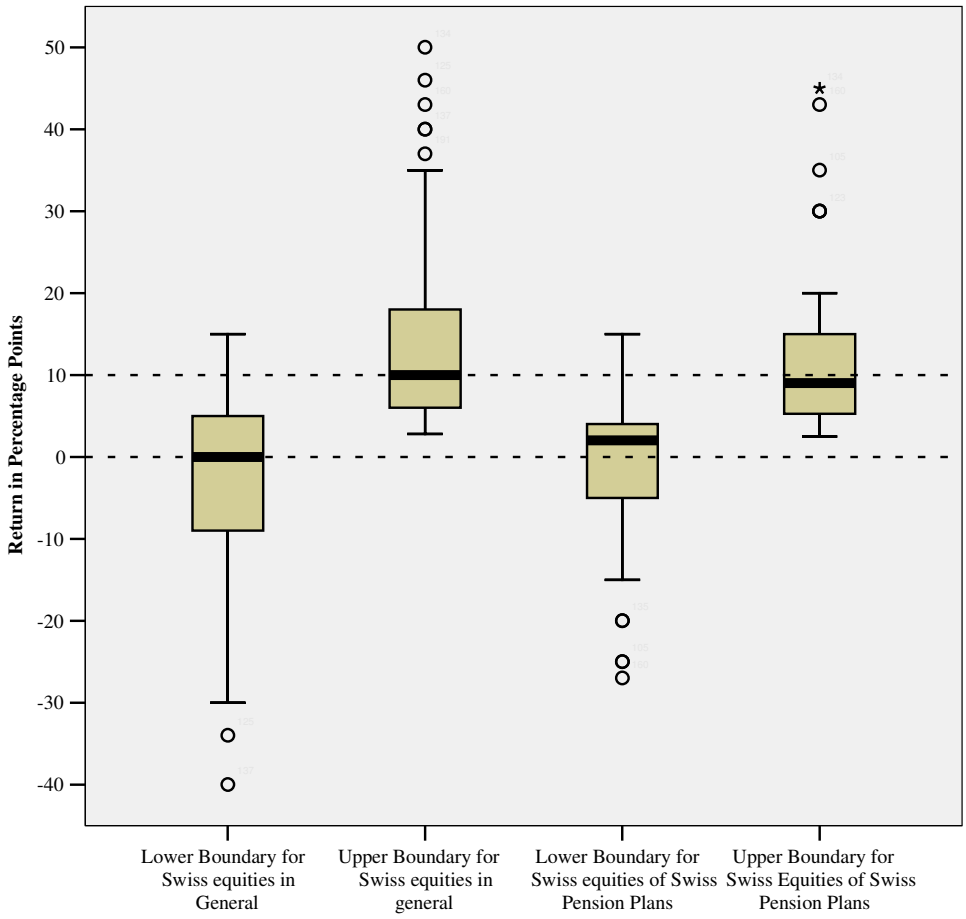
* 10% significance level
** 5% significance level
*** 1% significance level

This section can be concluded by saying that decision-makers of Swiss pension plans are significantly prone to the better-than-average-effect when evaluating the chances of future success for their current managers and their own pension plans. But they do not believe that they personally can forecast future returns of different asset classes significantly better than their colleagues from other Swiss pension plans.

Besides the participants' choice of answers on a Likert scale in the questionnaire there is another evidence of susceptibility to a better-than-average-effect in the data from the forecasting task. It is related to the participants' formulation of 90% confidence intervals for return forecasts in different asset classes. The participants in the sample forecast more attractive risk-return-characteristics for domestic equities at Swiss pension plans compared to domestic equities in general as can be seen in figure 3.2.

Figure 3-2: Return forecasts for Swiss equities

In figure 3.2 the distributions of the participants' subjective 90% confidence intervals for Swiss equities return forecasts are shown as boxplot diagramms. The Y-Axis shows the return in percentage points for the forecasts for upper and lower boundaries. The solid black line in each boxplot indicates the median and the box itself indicates the second and third quartile of the distribution. Bullets (stars) reflect outliers that are more than 1.5 (3) times above or below the top and bottom end of each box. The dotted lines facilitate to see the differences between forecasts for Swiss equities as an asset class in general and forecasts for Swiss equities at Swiss pension plans.



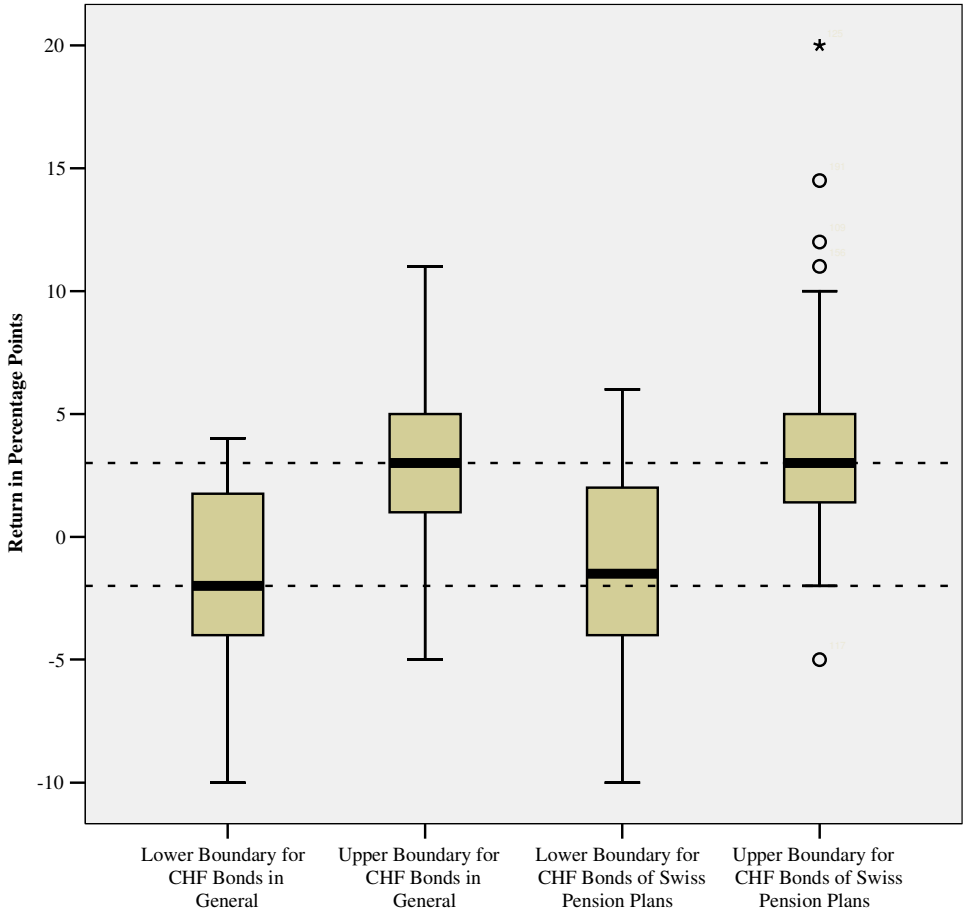
The differences of the upper and lower boundaries are significant at the 5% level in a Wilcoxon signed rank test.

Figure 3.2 summarizes the participants' lower and upper boundaries in the return forecasting task for the asset class domestic equities in general and in Swiss pension plans with four boxplots. The participants expect Swiss pension plans to reduce downside risk

of domestic equities but also give up some upside potential because the lower (higher) boundaries of their confidence intervals for pension plans' exposure to domestic equities are higher (lower) than for domestic equities in general. The fact that the median boundaries of the confidence intervals are narrower for the pension plans' exposure to domestic equities than for domestic equities in general indicates that the participants expect pension plans to manage this asset class with less volatility than a broad market index. Pearson and Tukey (1965) numerically show that dividing a 90% quantile (i.e. subtracting the 5% quantile from the 95% quantile) by 3.25 approximates the standard deviation of a distribution. Therefore narrower confidence intervals can be related to lower expected volatilities. To test the level of significance across boundary differences a non-parametric Wilcoxon signed rank test is applied. It shows that the differences between the lower and the upper differences are significant at the 5% level. Parametric tests would not fit the data because the participants' forecasts are far away from being normaly distributed and outliers might bias the results. Figure 3.3 contains the same analysis for the asset class CHF bonds. It shows that the decision-makers of Swiss pension plans on aggregate expect to match the downside risk of CHF bonds in general as there are no significant differences between the participants' lower boundaries. However the participants expect to have a higher upside potential in CHF bonds as the upper boundaries are significantly higher for CHF bonds in Swiss pension plans than in general. Figure 3.4 shows the lower and upper boundaries for the forecasted returns of the participants' own pension plans versus an average Swiss pension plan. The participants forecast a higher upside potential for the own pension plan and a lower downside risk and both effects are significant according to Wilcoxon signed rank test. Those results are in line with the participant's propensity to the better-than-average-effect about the future success of their managers and their pension plans.

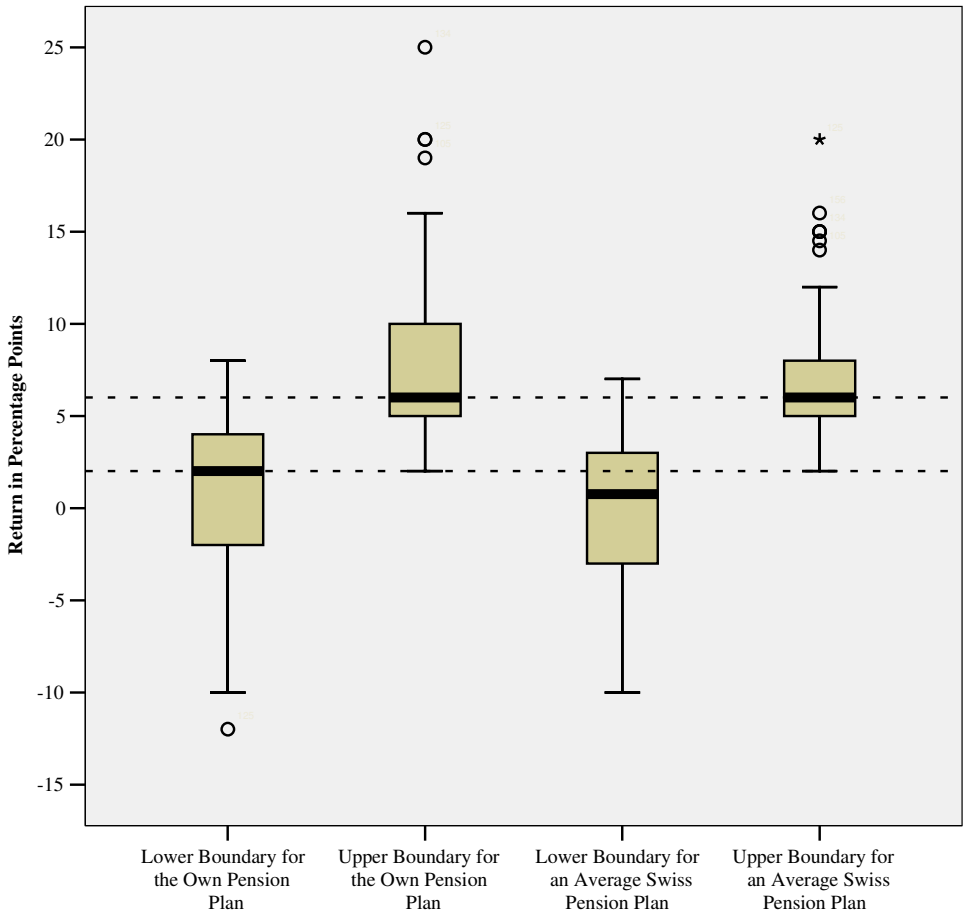
Figure 3-3: Return forecasts for CHF bonds

In figure 3.3 the distributions of the participants' subjective 90% confidence intervals for CHF bonds return forecasts are shown as boxplot diagramms. The Y-Axis shows the return in percentage points for the forecasts for upper and lower boundaries. The solid black line in each boxplot indicates the median and the box itself indicates the second and third quartile of the distribution. Bullets (stars) reflect outliers that are more than 1.5 (3) times above or below the top and bottom end of each box. The dotted lines facilitate to see the differences between forecasts for CHF bonds as an asset class in general and forecasts for CHF bonds at Swiss pension plans.



The differences of the lower boundaries are significant at the 5% level in a Wilcoxon signed rank test.

Figure 3-4: Return forecasts for Swiss pension plans
In figure 3.4 the distributions of the participants' subjective 90% confidence intervals for return forecasts for Swiss pension plans are shown as boxplot diagramms. The Y-Axis shows the return in percentage points for the forecasts for upper and lower boundaries. The solid black line in each boxplot indicates the median and the box itself indicates the second and third quartile of the distribution. Bullets (stars) reflect outliers that are more than 1.5 (3) times above or below the top and bottom end of each box. The dotted lines facilitate to see the differences between forecasts for an average Swiss pension plan and forecasts for the own pension plan.



The differences of the upper and lower boundaries are significant at the 5% level in a Wilcoxon signed rank test.

3.6 Discussion and conclusion

Some Swiss pension plans delivered extraordinary good performances in equities over the last couple of years but a majority of the Swiss pension plans that participate volun-

tarily in well established performance surveys delivered equity performances below market indices on an absolute as well as on a risk-adjusted basis even gross of fees. This does not imply that active management cannot add value for Swiss pension plans's equity portfolios but it shows how difficult the selection and the maintenance of successful active managers are. Many asset managers point out that past performance is no indication for future performance so it cannot be ruled out that the managers in this sample will outperform broad market indices in the future. But in order to meet expectations future results must be much better than historical results.

To put this evidence into perspective we have to bear in mind that we cannot directly link a participant's answers in the questionnaire to the investments of his/her pension plan as we do not know how big the influence of one decision-maker on the whole pension plan investment is. In other words, just because one decision-maker expresses biased expectations does not imply that his pension plan will fail with active management. Nevertheless it is puzzling that most of the decision-makers in the sample are highly convinced about the abilities of their current managers, their own manager selection skills and the chances for future success of their pension plan despite the observable track record. It's even more puzzling when you consider that a comparison with the average is very easy on financial markets as market indices are freely available and often discussed in the media. So where is this gap between biased expectations and realized performances in equities coming from? As this paper suggest, the better-than-average-effect is one explanation. A few arguments are now discussed in the context of Swiss pension plans.

At first glance one might argue that the pension plans in this sample might not be aware of their disappointing relative performance versus broad market indices. However this argument is weak because the pension plans in this sample are widely regarded as leading pension plans in Switzerland and they participate voluntarily in popular surveys of Swiss pension plans. They receive the results of the Lusenti and the ASIP surveys automatically as they are participants of those studies so it seems unlikely that they never look that the results of those surveys. So the assumption that the decision-makers of Swiss pension plans are aware of poor relative median performance of Swiss pension plans versus market indices seems safe. Of course, for all pension plans who have not yet implemented a proper measurement of their performances in different asset classes

relative to appropriate market indices the recommendation clearly is to start sooner rather than later.

Another very simple explanation could be that if the decision-makers of Swiss pension plans kept on believing to have above average managers and above average selections skills they might still be convinced to outperform in the long run despite the disappointing results so far. Indeed it takes a lot of time and data to be able to correctly differentiate between a managers' level of skill and pure luck as Waring and Siegel (2003) explain. If that was the case the participants in our sample will not put a lot of weight on their recent performance but more on their perception about their chances of future performance. Further research needs to be done in order to understand how much an individual weighs knowledge from the past versus hope for the future.

Less simplistic arguments can be found in the domain of the decision-makers' incentives to express a highly confident view about current managers in the own pension plan and about the own managers selection skills. Camerer and Lovallo (1999) offer the so called reference-group-neglect as an explanation for why people could seem to be prone to the better-than-average-effect. If the participants in a sample do not pay attention to the definition of the peer group – in this paper the Swiss pension plans that report to the ASIP survey as was clearly stated – they might compare themselves with an inappropriate peer group – in our case maybe with all pension plans in Switzerland. Because of the self-selecting nature of the ASIP survey and the required use of a global custodian for the data provision it is possible that the average performance of the pension plans in the ASIP survey is above the unobservable average of all pension plans in Switzerland. So the benchmark to beat in our sample might be higher than a benchmark that consists of all Swiss pension plans and the reference-group-neglect offers an explanation for the high proportion of participants that demonstrate the better-than-average-effect.

Furthermore it might be the case that some decision-makers in our sample are more concerned with being above average in the Swiss pension plan universe than with the relative performance versus a market index (this is the case for UK pension plans according to a study by Blake, Lehman and Timmermann (2002)). However this argument is no excuse for the poor relative performance versus market indices because the goal of active management in traditional asset classes like equities and bonds is usually to outperform market indices and not pension plan peer groups.

Lakonishok, Shleifer and Vishny (1992) present other incentives for pension plan managers to apply active management despite a lack of success in the past. One reason could be job security. Active management requires the decision-makers to select active managers. This is arguably a more complex and interesting process when compared to the selection of indexers, which tends to prioritize the factor price. This study suggests that some decision-makers may fear that their ability to add value is diluted when too high a proportion of their plan's assets are indexed. Decision-makers have strong incentives to demonstrate confidence about their active managers and their own manager selection skills going forward, and to ignore disappointing performance relative to market indices in the past to defend their employment.

Another reason could be that the performance of an active manager might not be the whole story behind an appointment to run portfolios for a pension plan. Lakonishok, Shleifer and Vishny (1992) call such a phenomenon "schmoozing". It describes the colorful manner that asset managers sometimes explain their stock decisions, hold hands with their clients and explain their absolute and relative performance ex post. With successful "schmoozing", active managers might be able to convince decision-makers at Swiss pension plans about their skills in spite of an unsuccessful track record. As outlined before, investment committees of Swiss pension plans do not only consist of investment professionals but also of delegates of the board of trustees who might be less familiar with daily portfolio management. And it seems plausible that an active manager with excellent schmoozing abilities is in a good position to be retained by a pension plan despite a relative underperformance versus market indices.

In line with this thinking another incentive for Swiss pension plans not to replace unsuccessful active managers can be seen in the linkage of other corporate services some captive active managers provide. Large asset managers typically affiliated to banking groups not only offer the sponsor of a pension plan active management services for the pension plan's assets but also services in the field of credit management, refinancing, mortgages, financial consulting, currency management and many other services to the sponsoring businesses. The decision-makers at the sponsor organization have incentives to retain active managers for the pension fund which also provide such other valuable services to the company. Especially in a defined contribution scheme where the risk of the retirement payments are borne by the employees a company can pay for many ser-

vices via management fees for pension plan assets and take those costs off the current account statements of the company. To express a strong belief in the future success of active managers might be a convenient way to hide the true motivation for keeping active managers with a disappointing relative performance. Further research is needed to analyze the incentives on a personal level and the effects of the regulatory framework with respect to the separation of the pension plan and the sponsor group.

In the case a pension plan employs in-house active managers the problem has also a personal aspect. It is likely to be much easier to hire and fire external active asset managers than in-house managers because the personal relationship with external people is usually less strong than with in-house employees. This might be a reason why the participants in our sample indicate that they are confident about the skill of their current internal active managers relative to the other pension plans' managers. Probably those answers reflect the view that there are no opportunities for a replacement and the current setup is still perceived as attractive despite the poor track record. Pretending to feel confident with a mediocre in-house active manager could probably be more convenient for an investment committee member than firing this colleague. Such qualitative arguments appear plausible at first glance but require more research to analyze the interdependent relationships in more depth and to prove or reject these hypotheses.

A completely different type of incentive for active management despite a bad track record is offered by Odean and Barber (2001). It might be the case that active management is perceived as more entertaining than indexing by the decision-makers of Swiss pension plans. As a proxy to measure the degree of entertainment it is also analysed whether there are any differences in the answering patterns between those participants who have a private asset management account and those who do not. An independent t test shows that the management of an own account has neither a significant influence on their proneness to the better-than-average-effect nor on the formulation of return forecasts. This is evidence that the degree being entertained by actively managing money is not related to each other in this sample.

After the discussion of potential arguments that explain the occurrence of a better-than-average-effect in the sample we finally turn to a recommendation about how to deal with it. The issue is theoretically addressed in a model by Thorley (1999). He compares the competition to generate an outperformance on financial markets to a basketball

freethrow shooting contest. Every participant can choose between throwing freethrows himself or not throwing and just getting the average score. The throwing corresponds to active management and the latter to indexing. It only makes sense to shoot the freethrows if a participant believes to be an above average freethrow shooter compared to all the other contestants in the game. If only children are playing on the court shooting freethrows as an adult might be a good opportunity to beat the average score. By contrast when on the court with the basketball dream team from the USA not shooting oneself but taking the average score is probably a more successful strategy. So the decision about implementing the strategic asset allocation with active managers should depend on how the overall level of competition and the own abilities compared to the competitors are perceived. The better-than-average-effect might bias the perception of personal abilities and skills relative to the other competitors. The historical performance of Swiss pension plans supports the view that not every Swiss pension plan is aware of its strengths and weaknesses relative to the other competitors on financial markets. So the model recommends to analyze in which asset classes the pension plan should have a true advantage versus most of the participants in an asset class and only apply active management in those asset classes.

3.7 Literature

ASIP. "ASIP Performance Comparison." ASIP, 2000-2006.

Berzins, J., and Trzcinka, C. "Performance and Survival of Institutional Money Managers." *Working Paper*, 2005.

BFS, B. f. S. "Pensionskassenstatistik." Bundesamt für Statistik BFS, Neuchatel, 2004.

Blake, D., Lehman, B., and Timmermann, A. "Asset Allocation Dynamics and Pension Fund Performance." *The Journal of Business,,* 72, (1999), pp. 429-461.

Blake, D., Lehman, B., and Timmermann, A. "Performance Clustering and incentives in the UK pension fund industry." *Journal of Asset Management,* 3, (2002), pp. 173-194.

Camerer, C., and Lovallo, D. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review,* 89, (1999), pp. 306-318.

Carhart, M. "On Persistence of Mutual Fund Performance." *Journal of Finance,* 52, (1997), pp. 57-82.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. "Measuring Mutual Fund Performance with Characteristic Benchmarks." *Journal of Finance,* 52, (1997), pp. 1035-1058.

De Long, B., Shleifer, A., Summers, L., and Waldman, R. "The Survival of Noise Traders in Financial Markets." *The Journal of Business,* 64, (1991), pp. 1-19.

Ennis, R. "The Structure of the Investment Management Industry: Revisiting the new Paradigm." *Financial Analysts' Journal,* 53, (1997), pp. 6-13.

Fama, E. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance,* 25, (1970), pp. 383-417.

Frauenlob, T. "Anlagephilosophien von Schweizer Pensionskassen." Haupt Verlag, Bern, 1998.

Glaser, M., and Weber, M. "Overconfidence and Trading Volume." *Geneva Risk and Insurance Review,* 32, (2007), pp. 1-36.

Graham, J., Campbell, H., and Huang, H. "Investor Competence, Trading Frequency and Home Bias." *NBER Working Paper*, W11426, 2006.

Grossman, S., and Stiglitz, J. "On the impossibility of informationally efficient markets." *American Economic Review,* 70, (1980), pp. 393-408.

Guiso, L., and Jappelli, T. "Information Acquisition, Overconfidence and Portfolio Performance." *CEPR Discussion Paper*, 5901, 2005.

Kyle, A., and Wang, A. "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?" *Journal of Finance,* 5, (1997), pp. 2073-2090.

Lakonishok, J., Shleifer, A., and Vishny, R. "The Structure and Performance of the Money Management Industry." *Brookings Papers on Economic Activity. Microeconomics,* (1992), pp. 339-391.

Lusenti, P. "Anlagen Schweizer Institutioneller Investoren: Umfrageergebnisse." Lusenti Partners, Nyon, 2003.

Lusenti, P. "Swiss Institutional Survey: Umfrageergebnisse 31.12.2004." Lusenti Partners, Nyon, 2005.

Lusenti, P. "Swiss Institutional Survey: Umfrageergebnisse 31.12.2005." Lusenti Partners, Nyon, 2006.

Lusenti, P. "Swiss Institutional Survey: Umfrageergebnisse 31.12.2006." Lusenti Partners, Nyon, 2007.

Malkiel, B. "A Random Walk down Wallstreet." W. W. Norton & Company, 2004.

Odean, T., and Barber, B. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance,* 55, (2000), pp. 773-806.

Odean, T., and Barber, B. "Boys will be boys: Gender, Overconfidence and Common Stock Investment." *The Quarterly Journal of Economics,* 116, (2001), pp. 261-292.

Pearson, E., and Tukey, J. "Approximate means and standard deviations based on distances between percentage points and frequency curves." *Biometrika,* 52, (1965), pp. 533-546.

Roll, R. "A Critique of the Asset Pricing Theory's tests. Part One: On past and potential testability of the Theory." *Journal of Financial Economics,* 4, (1977), pp. 129-176.

Sharpe, W. "Capital Asset Prices: A Theory of Market Equilibrium." *Journal of Finance,* 4, (1964), pp. 425-442.

Sharpe, W. "The Arithmetic of Active Management." *Financial Analysts' Journal,* 47, (1991), pp. 7-9.

Siegel, L., Kroner, K., and Clifford, S. "The Greatest Return Stories Ever Told." *The Journal of Investing,* Summer 2001, (2001), pp. 91-102.

SNB. "Schweizerische Nationalbank - Finanzierungsrechnung der Schweiz 2004." Swiss National Bank, 2006.

Stanko, D. "Performance Evaluation of Public Pension Funds: The Reformed Pension System in Poland." *The Pensions Institute*, PI-0308, 2003.

Svenson, O. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica,* 47, (1981), pp. 143-148.

Swisscanto. "Schweizer Pensionskassen 2006." Swisscanto, 2006.

Taylor, S., and Brown, J. "Illusion and well-being: A social psychological perspective on mental health." *Psychological Bulleting,* 103, (1988), pp. 193-210.

Taylor, S., and Brown, J. "Positive Illusions and Well-being revisited - Separating Fact from Fiction." *Psychological Bulleting,* 116, (1994), pp. 21-27.

Thorley, S. "The Inefficient Markets Argument for Passive Investing." 1999.

Tonks, I. "Fund Manager Performance of Segregated UK Pension Funds." *CMPO Working Paper*, 01/33, 2001.

Waring, B., and Siegel, L. "The Dimensions of Active Management." *Journal of Portfolio Management,* 29, (2003), pp. 35-51.

3.8 Appendix I

The questionnaires provided in this section are translated from the German original questionnaires.

The questions about the own ability to forecast returns in the year 2006 for different asset classes relative to the other decision-makers of Swiss pension plans in this sample have been worded the following way:

"Please evaluate the accuracy of your personal forecast relative to the forecasts of the other decision-makers of Swiss pension plans in this sample. Please tick the corresponding box."

Clearly below Average	Below Average	Rather below Average	Average	Rather above Average	Above Average	Clearly above Average
O	O	O	O	O	O	O

For the questions about the future chances of internal and external managers as well as the own pension plan the same Likert scale form 1 to 7 has been used. The questions have been worded the following way:

"How do you evaluate your pension plan's chances to find (internal or external) active managers who will perform above average in the future?"

"How do you evaluate your pension plan's chances to outperform the average performance of the other Swiss pension plans who participated in this survey on a risk-adjusted basis?"

"How do you evaluate your internal (external) managers' chances to outperform the internal (external) managers of the other Swiss pension plans who participated in this survey?"

4. The Performance of Groups and Individuals in Financial Decision-Making

Abstract

On financial markets many investment decisions are taken by groups and not by individuals. The evidence, however, whether groups perform better than individuals, is ambiguous. We analyze the portfolios of groups and individuals in an asset allocation task on an experimental market. We find that groups on average outperform individuals, i.e. achieve higher Sharpe Ratios but the difference is not significant. Furthermore, stochastic groups built of randomly chosen individuals perform slightly better than the real groups on average. But an analysis of our group sample average loses sight of the large performance discrepancies across groups. An important determinant of the success of the groups in our experiment is the degree of information exchange between group members as a higher level is linked to a significantly better performance.

4.1 Introduction

Today many decisions in the domain of financial markets are taken by groups and not by individuals. For example asset management firms often appoint management teams instead of single managers and institutional investors like pension plans elect boards of trustees to define the strategic asset allocation and investment committees to decide about tactical allocations. However, economic research so far has not put a lot of attention on differences between the outcome of individual and group decisions. According to Kocher and Sutter (2005) one reason might be that many economic models assume the decision-maker to choose equilibrium (or optimal) actions. In such cases it does not matter whether the decision-maker is an individual or a group. But the question is whether it matters in practice when real people take decisions.

The main research contribution of this paper is to analyze the performance of a sample of groups and a sample of individuals in an asset allocation task. On an experimental market a sample of individuals and a sample of groups were asked to build portfolios of 3 risky assets and 1 riskfree asset. Subjects were paid according to the Sharpe Ratio they achieved during the experiment. We analyze whether groups take economically more efficient decisions than individuals and find that on average groups do not significantly but only marginally outperform individuals.

A focus solely on the average performance of groups, however, disregards the heterogeneity in our sample. There are large performance discrepancies between the best and the worst groups and they seem to be related to different processes groups apply to agree on an asset allocation in our experiment. Recordings of the discussions within each group during the experiment allow us to judge how much information the group members exchanged and how they chose their asset allocation. In line with the hidden profile thesis of Stasser and Titus (1985) the success of a group in the experiment is positively related to the level of information exchange across group members. So groups do not generally take economically more efficient decisions than individuals but have the potential to do so if the group members exchange their personal information and their ideas.

The rest of the paper is organized as follows. Section II provides an overview of related research about group decision making and section III describes the hypotheses that are tested. Section IV outlines the design of the experiment and describes the data. Section

V contains the results of our experiment and section VI discusses our findings and concludes.

4.2 Related research

The existing literature provides no generalized theoretical framework about the differences of judgments between groups and individuals in economic decision-making and the evidence from empirical studies is mixed. However, research in psychology and economics offers various insights into the decision-making processes of groups and their performance relative to individuals. This section first discusses an important distinction between different types of experimental tasks namely judgmental and intellective tasks. Then it lists results from previous studies about economic decision-making of groups versus individuals in psychology and in economics to provide an overview of the heterogeneity of the results. Finally the focus is on two psychological aspects which address possible outcomes of group decisions relative to individual decisions in general.

The psychological literature distinguishes between judgmental tasks and intellective tasks as outlined by Davis (1992). The former are characterized by not having an undisputable optimal solution whereas the latter typically have a self-confirming and easily demonstrable optimal solution. For example tasks in which subjective risk preferences are measured can be categorized as judgmental. In contrast logic problems like the Tower of Hanoi are intellective problems as Cooper and Kagel (2005) explain. The asset allocation task in our experiment is an intellective one. First, because it has an optimal solution - the tangency portfolio which theoretically could be calculated by each participant who is familiar with portfolio optimization – and second because the participants receive feedback about their performance during each round of the experiment so there is the potential to find evidence about the success of a certain strategy. But we acknowledge that given the information we have provided in the experiment the calculation of the tangency portfolio is extremely demanding without a computer and very difficult to demonstrate ex ante.

Representative for the wide variety of results of judgmental tasks in psychology Kerr, Kramer and MacCoun (1996) provide a survey of empirical studies about decision-making of groups versus individuals. They point out that many differences with respect to the methodology, the group framing and the task processing make a meta-study al-

most impossible. The authors further note that the group decision-making process (for example majority voting, equilibrating or unanimous voting) is highly influential on the outcome as it can attenuate, amplify or reproduce the judgment biases of individuals.

The evidence in psychological research on intellective tasks is less multisided. Before 1955 it seemed clear to both research and common sense that groups outperform individuals in intellective tasks (see Shaw (1932) for a classic experiment). But then Lorge and Solomon (1955) re-examined Shaw's classic study and introduced a new standard called "truth-wins-standard" to measure the superiority of group versus individuals. It states that groups will find a correct solution to an intellective task if at least one group member proposes it. The larger the size of the group is the larger are the probabilities for synergies and the chances to solve the problem. The re-examination of Shaw's and other data revealed that groups rarely outperform individuals, sometimes match and usually fall below the truth-wins-standard as Davis (1992) summarizes. So a meaningful judgment of the performance of groups requires not only a comparison with a sample of individuals but also a measurement relative to a truth-wins-standard.

The debate about the performance of groups relative to individuals continues as more recent evidence from psychological studies of intellective tasks is available. As far as rationality in the decision-making process is concerned Yaniv and Bornstein (1998) show that groups tend to take more rational decisions than individuals because the groups in their sample tendered lower offers and accepted lower bids in a one-shot ultimatum game. According to a study by Laughlin, Bonner and Altermatt (1998) about formulating hypotheses in an intellective task with playing cards, groups of four participants perform at the level of the best individual and therefore significantly above individuals' average. They argue that the group performance increases with task complexity and information variety as groups can better handle such issues. Rockenbach, Sadrieh and Mathauschek (2006) report that groups achieve similar levels of expected payoffs with significantly lower risk than individuals when choosing between different lotteries. Bornstein, Kugler and Ziegelmeyer (2004) report that groups exit a centipede game significantly earlier than individuals which is more rational from a game theory perspective. Concluding, the authors point out that groups and individuals make different decisions in strategic games and that in a majority of the cases group decisions are closer to the "rational" solution. Counter evidence is offered by Cason and Mui (1997) who

found that groups make more generous allocations in a dictator game than individuals and were thus further away from the game-theoretic prediction. Kocher and Sutter (2005) report that groups learn faster than individuals in a beauty-contest game on an experimental market but they point out that groups are neither smarter nor better decision-makers than individuals per se.

As in psychology the bulk of economic evidence about decision-making in intellective tasks so far indicates a tendency for groups to take more rational decisions than individuals. Cooper and Kagel (2005) report that groups of two persons tend to take more strategic decisions than individuals in a signaling game even if the truth-wins-standard is taken into account. Blinder and Morgan (2005) report that groups tend to take better decisions than individuals in a monetary policy game. But they find no significant differences between majority decisions and unanimous decisions as far as groups' performances are concerned. Lombardelli, Proudman and Talbot (2005) report in a similar study that groups achieve significantly better results than individuals in a policy making game on an experimental market and that groups even outperform the best individuals. They argue that the worst individual decisions are averaged out in the group and that the members of the groups learn from each other. But they also note in their sample that the opportunity to discuss in one part of the sample did not enhance the performance relative to the other part of the sample in which groups were not allowed to discuss.

The results of empirical studies on real financial markets are in line with those from experimental markets as no stable differences between the performance of groups and individuals have been found so far. Barber and Odean (2000) report from a sample of 166 US investment clubs out of 78'000 trading accounts that the clubs lag the performance of the S&P 500 Index by 3.7% and the performance of individual investors in this sample by 2.3% after costs. However, they note that clubs' portfolios show on average lower monthly volatilities than individuals' portfolios - due to a higher degree of portfolio diversification - and therefore achieve a higher Sharpe Ratio. Baer, Kempf and Ruenzi (2005) find in their sample of mutual funds that funds managed by teams show lower levels of unsystematic risk. Such funds also change their risk to a lesser extent as a response to prior performance than funds managed by individuals. However, there is no significant difference between the performances of team managed funds and individually managed funds overall.

After this overview about the heterogeneity of decision-making of groups relative to individuals we now turn to two psychological aspects about how the decision-making process in a group might impact the outcome. Those aspects serve as explanations for the heterogeneity of the results in our own group sample and also help to interpret the evidence of our asset allocation experiment. The first aspect addresses the level of information exchange across group members as one requirement of successful group decisions. The second focuses on the outcome of a group decision when individual positions and views are exchanged across group members.

Stasser and Titus (1985) offer one explanation for why group decisions might not be fully rational. They report that groups tend not to use all potentially available information in their decision making process in a political voting game but rather focus on information that is both shared by everybody in the group and in line with the general thinking of the group. In the authors' experiment every group member received different but slightly overlapping information about candidates in the voting game. To find the best candidate the group members had to share all their available information but few groups did so and because of this lack of information exchange opted for the wrong candidate. This effect is referred to as "hidden profiles" and can lead to bad decisions despite the fact that the group as a whole had all necessary information to take a rational decision. So the recommendation for groups is to share all available information. Janis (1982) offers supportive evidence. Based on studies about American foreign policy decisions since 1940 he concludes that a process for successful group decisions is characterized, among other things, by the rational weighting of possible options in the light of all available evidence. Those results imply that there might be differences with respect to the performance of groups in our asset allocation experiment because of different levels of information exchange across group members.

When analyzing the impact of information exchange across group members with respect to performance, we need to assess how the information is processed in a group. The evidence as to whether groups moderate or increase individual positions in the decision-making process is ambiguous. Until the Sixties it was widely believed that group decisions are more or less an average reflection of individual beliefs. Today this view is supported by the theoretical work of Sah and Stiglitz (1986) who argue that group decisions are less extreme and less volatile than individual decisions or by Moscovici (1985)

who presents empirical evidence that group decisions may shift individual attitudes towards more moderate positions. In contrast Stoner (1961) counters those predictions by describing the so called “risky shift” effect. It refers to the observation that groups do not moderate extreme positions of individuals after a discussion of the issue in a group but rather emphasize those extreme positions even more. Moscovici and Zavalloni (1969) generalized this line of thinking and introduced the “group polarization” effect. They argue that group polarization does not necessarily lead to risky shifts but can also lead to “cautios shifts”. So following an information exchange between group members, a group's decision will be a more extreme version of individual's preferred action but it is not possible to say ex ante whether it will be riskier or more cautious.

4.3 Hypotheses

We now turn to the hypotheses that will be tested in this paper. In general our null hypotheses postulate that there are no differences between groups and individuals with respect to portfolio construction in our asset allocation experiment. The reason is, as outlined in the section above, that the evidence about the decision-making quality of groups in judgmental and in intellective tasks relative to individuals is mixed. We not only compare the performance of individuals and groups in our experiment but also judge the quality of those allocation decisions by the groups in our sample. Therefore, we apply 3 different types of hypotheses (1.1, 1.2 and 1.3) that address the performances in our experiment. In addition we formulate hypotheses 2 and 3 to analyze the level of risk our participants are willing to take. Last but not least hypothesis 4 is focused on differences regarding the level of information exchange within the sample of groups.

Hypothesis 1.1: On average the asset allocations of groups are as efficient as the asset allocations of individuals.

Hypothesis 1.2: On average the asset allocations of groups are as efficient as asset allocations which are calculated by averaging out the asset allocations of 3 randomly chosen individuals.

Hypothesis 1.3: On average the asset allocations of groups are less efficient than asset allocations which represent the best single asset allocation out of 3 randomly chosen individuals.

Hypothesis 2: On average the portfolios of groups are as volatile as the portfolios of individuals.

Hypothesis 3: On average groups shift the weightings of each asset in their asset allocations as much as individuals from one round to the next.

Hypothesis 4: Groups with a higher level of information exchange between group members outperform groups with a lower level.

We will rely on Mann-Whitney tests to test the significance of our hypotheses. Mann-Whitney tests are nonparametric which is important for the analysis because the distributions of the Sharpe Ratios and the asset weightings of individuals and groups are very different from a normal distribution. The experimental design to address our hypotheses is explained in the next section.

4.4 Experiment

4.4.1 Experimental design

Our experiment consists of an asset allocation task with 3 risky assets and 1 riskfree asset. To test our hypotheses about the groups' and individuals' asset allocation efficiency we need a framework to measure the efficiency and the degree of risk. According to the arguments from Bossaerts and Plott (2002) a CAPM framework makes sense to set up such an experimental market and to compare the participants' Sharpe Ratios. They argue that the empirical difficulties to test the CAPM do not rule out its usefulness in an experimental market because the many divergences from reality can be controlled. The Sharpe Ratio of a portfolio is a well established measure to compare the efficiency of different portfolios and so we can rank the participants' portfolios according to their Sharpe Ratios and compare groups' and individuals' achievements. One advantage of applying the Sharpe Ratio as a measure for portfolio efficiency compared to the return is that the Sharpe Ratio is not subject to the participants' degree of risk aversion. So the target function in the experiment is the same for all participants regardless of their risk preferences and this allows an unbiased comparison of the performances. To test hypothesis 2 we use only the denominator of the Sharpe Ratio which is the volatility of a portfolio. To examine hypothesis 3 we analyze the weightings of each asset in each round of the asset allocations.

Instead of informing the participants directly about the expected returns and the volatilities of the assets we have defined 4 different but equally probable states of nature (Bossaerts and Plott (2002)) with distinct returns for all of our assets. Table 4.1a contains a summary of the states of nature and the returns of each asset in those states. A detailed provision of information about the asset's volatilities and correlations would have allowed the participants with an economic background to easily calculate the optimal asset allocation. However, to make our experiment as realistic as possible, we did not provide such information because, in reality, the correct covariance matrix is also unknown on an ex ante basis.

To keep our asset allocation experiment in line with the characteristics of real financial assets we have defined risk premia for the assets that increase with the volatility of the returns. Also the returns have been defined to have low correlations between the assets in the different states of nature. The goal in our experiment is to maximize this Sharpe Ratio, which is similar to finding the tangency portfolio. To make this search challenging we have defined very different weights for each risky asset as the tangency portfolio weights in table 4.1b show. This assures that it is hard to find the optimal portfolio by accident and that the odds are small that averaging out individual allocations across group members lead automatically to the optimal solution. Also the optimal solution is far away from either an allocation of 100% in one asset or an equally weighted allocation.[14] Table 4.1b also contains the excess returns of the risky assets and the covariance matrix that is needed to solve our asset allocation problem to find the maximum Sharpe Ratio. The participants in our experiment did not receive any information from table 4.1b.

[14] An alternative approach would have been to draw random states of nature. However, in our pilot experiment we realized that the effect of random states increases the difficulty of the asset allocation task considerably because the feedback is blurred due to its randomness. As a consequence it seemed as if the participants were not able to learn from previous rounds. We acknowledge that in our framework the participants do not get a fuzzy feedback about the efficiency of their allocation which would be the case in real world markets. However the lack of randomness does not affect our research goal which is to analyze performance differences between individuals and groups and we did not want to threaten this study by relying on random results.

Table 4-1: Parameters of the experiment

The table 4.1a lists the different assets (A, B, C and D) in the experiment and their returns in the 4 different states of nature in the experiment. Every participant only received table 4.1a. Table 4.1b contains the excess returns of each asset above the risk-free rate, the covariance matrix resulting from the state-returns in table 4.1a and the weights of each asset in the tangency portfolio as well as the resulting maximum Sharpe Ratio achievable in every round. Only the relative weightings of the risky assets A, B and C are relevant for the calculation of the maximum Sharpe Ratio. For example and allocation of 3% in A, 33.5% in B, 13.5% in C and 50% in D respectively would have also returned the maximum Sharpe Ratio.

Table 4.1a

Szenario	Probability	Return Asset A	Return Asset B	Return Asset C	Return Asset D
I	0.25	15%	4%	0%	1%
II	0.25	-2%	5%	5%	1%
III	0.25	8%	1%	9%	1%
IV	0.25	-4%	0%	-1%	1%

Table 4.1b

		Asset A	Asset B	Asset C	Asset D
Excess Return		3.25%	1.50%	2.25%	0%
Covariance Matrix		Asset A	Asset B	Asset C	
	Asset A	0.0059	0.0004	0.0003	
	Asset B	0.0004	0.0004	0	
	Asset C	0.0003	0	0.0016	
Tangency Portfolio Weights		6%	67%	27%	0%
Sharpe Ratio of Tangency Portfolio: 0.923					

15 rounds, preceded by a test run, were played in the experiment to give the participants a chance to learn from results of previous rounds. In each of those 15 rounds the participants had to choose an asset allocation. After every round each individual and each group was informed about the mean return, the volatility and the Sharpe Ratio of their portfolio in that round as well as for the overall experiment. In addition every participant received information about the highest Sharpe Ratio a participant achieved in each round of the experiment. This feedback allowed the participants to better judge their own allocation. Such information also fits reality because the performance of (successful) competitors on financial markets is usually observable.

Hidden profiles were dampened in our experimental design to a certain extent because we provided everybody with the same set of information as every participant received

the same instructions (see appendix I for a translation of the original German instructions). Given this set of information every participant was in a position to find the optimal solution herself and this substantially reduces the issues with hidden profiles. Defining characteristics of a hidden profile according to Schulz-Hardt and Greitmeyer (2003) do not exist in our experimental design. However, the groups still face a special sort of a hidden profile effect. If the group members do not share their personal opinions and ideas about how to weight the different assets they might agree on a bad allocation simply because no group member could solve the problem but allows his personal view to prevail. The impact of the level of information exchange on performance is what we want to test in hypothesis 4 and so the discussions of each group have been recorded to collect data about the level of information exchange within our groups. The recording started with the distribution of the instructions of the experiment and it ended with the announcement of the results in the last round of the experiment.[15]

To set appropriate incentives the participants were paid according to their average Sharpe Ratio across all rounds of the experiment. The average payout per participant was 38.5 CHF. Also a bonus for the winning group in each group experiment and for the best four individuals in each individual experiment was paid. All participants were informed about the payoff structure via the instructions.

4.4.2 Data

The computerized experiments were conducted between June 15th, 2007 and July 11th, 2007 at the University of Zurich.[16] In total 93 students participated in our experiment, 43 female and 50 male, with ages ranging from 19 to 32. 16 study economics at the University of Zurich, 27 were other students from the University of Zurich and 50 were students from the Swiss Federal Institute of Technology (ETHZ). The participants have been recruited using the recruitment system ORSEE (see Greiner (2004)). 45 students participated in the experiment as individuals and 48 as part of a group of 3 people. In order to rule out any biases in the groups we constructed the groups randomly and neither gender nor education or any other personal characteristics were taken into account

[15] The recordings (in Swiss German) are available from the authors on request.

[16] The experiment was programmed and conducted with the experimental software "z-tree" (see Fischbacher, Urs, 1999, z-tree. Toolbox for Readymade Economic Experiments, IEW Working Paper, University of Zurich.).

when forming the groups. All participants drew cards with different letters and the persons with the same letters formed a group.

The experiments for the individuals were conducted in the laboratory of the Institute for Empirical Research in Economics and the group experiments in the Swiss Banking Institute at the University of Zurich. Two experiments with 24 and 21 participants were conducted to collect data about the individuals and four experiments with 5 (4 on one occasion) groups were conducted to collect the group data. There are no significant differences between individual participants and group members with respect to personal factors like education, age, gender and personal interest in financial topics. Therefore, differences between the performance of individuals and groups in the experiments cannot be attributed to personal characteristics of the participants.

To make sure that every participant fully understood the tasks in the experiment questionnaire I was distributed and the experiment did not start until each participant in each experiment successfully completed questionnaire I. The completion of questionnaire I took roughly 20 minutes in each experiment. To gather more information about the participants' experience as a group member or as an individual questionnaire II was distributed after the last round of the experiment. Appendix II contains a translation of both questionnaires from the German original.

4.5 Results

4.5.1 Performance comparison between individuals and groups

To judge the portfolio efficiency of individuals and groups (hypothesis 1.1) we compare the Sharpe Ratios and check if there is a significant difference in each round and across the whole experiment.[17] To calculate the Sharpe Ratio of each participant's portfolio in each round we relied on the theoretical distribution according to the states and returns defined in table 4.1a.

[17] 3 groups have been excluded from the analysis because some group members already completed the experiment as individuals. Their experience as individuals positively biased the overall results of those groups because they already knew the optimal allocation. As we reported the best Sharpe Ratio of an experiment to all participants it might be the case that having biased groups also biased the whole experiment. A Mann-Whitney test confirmed, however, that there is no significant difference at the 10% level in every but one round between group experiment sessions with and without exclusions. So the exclusion to those groups does not affect the results of all the other groups in any way.

Figure 4-1: Sharpe Ratios in the experiment

Figure 4.1a shows the average Sharpe Ratio per round for individuals, groups and both types of stochastic groups; average stochastic groups and best-in-group stochastic groups.

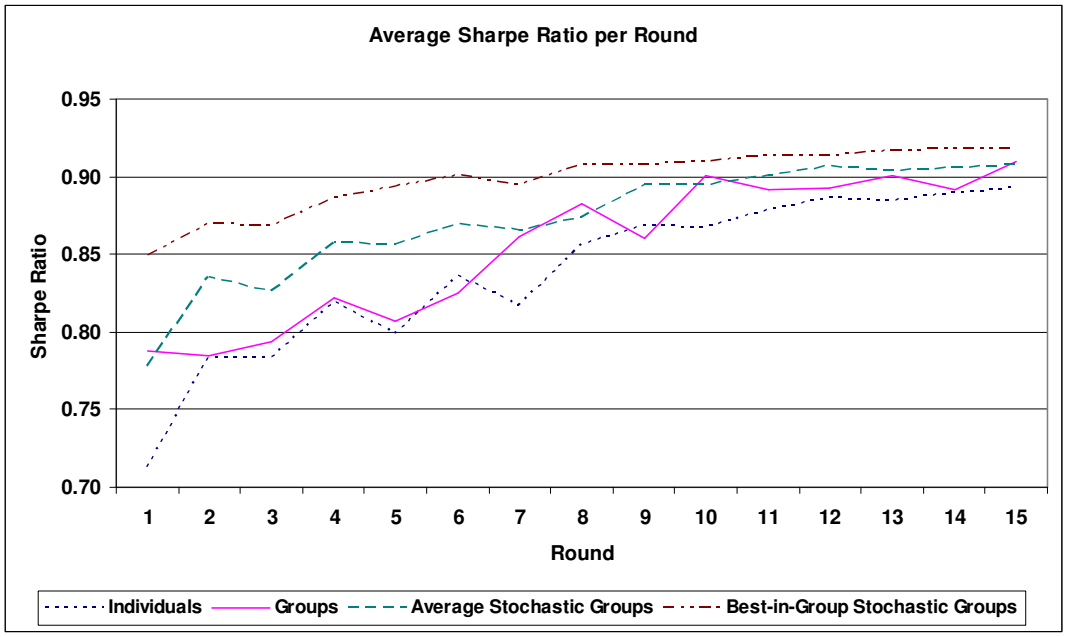


Figure 4.1b shows the median Sharpe Ratio per round for individuals, groups and both types of stochastic groups; average stochastic groups and best-in-group stochastic groups.

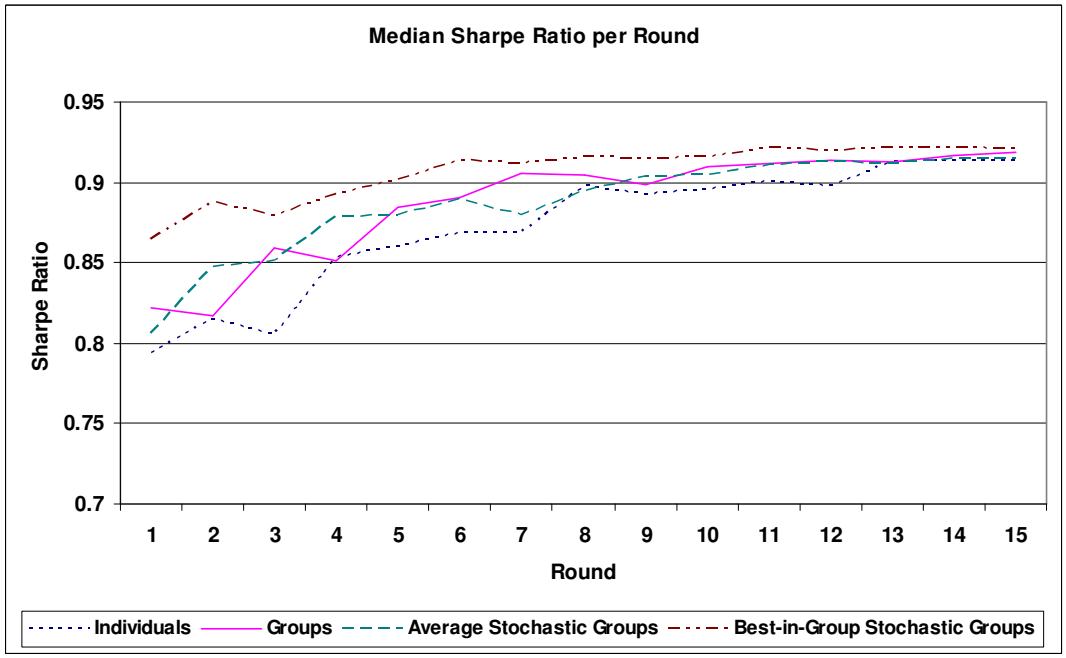


Figure 4.1a shows the average and figure 4.1b the median Sharpe Ratio for individuals and groups per round in our asset allocation experiment. The maximum Sharpe Ratio to be achieved in each round of our experiment is 0.923 (table 4.1b). It can be seen that groups achieve better Sharpe Ratios on average (and in the median) than individuals in 13 (15) rounds.[18] A Mann-Whitney test, however, reveals that groups do not significantly outperform individuals, as the difference between the achieved Sharpe Ratio of groups and individuals is not significant at the 5% level in any round. The Z scores of Mann-Whitney tests across all 15 rounds of the experiment range from -1.46 to -0.16 and the average is only -0.86 so we cannot reject hypothesis 1.1.

Furthermore, figures 4.1a and 4.1b reveal that both groups and individuals learn on average because the average and median Sharpe Ratios are more often increasing than decreasing over the 15 rounds in the experiment. There is no indication that one sample learnt faster than the other because the evolving trends of groups and individuals during the 15 rounds in the experiment were very similar. For individuals (groups) we observed 10 (9) increases on average and 11 (9) in the median. Unsurprisingly the learning effect was larger in the first half of the experiment than in the second half, and there is no single group and only one individual who performed better in the first round than in the last round.

An equal weighting of the risky assets in our experiments yields a Sharpe Ratio of 0.724. In figure 4.1a we see that groups on average achieved higher Sharpe Ratios already in the first round whereas individuals slightly underperformed an equally weighted allocation. But from the second round on both samples perform on average significantly better with their allocation according to a one sample t test. So, on average, individuals as well as the groups figured out very quickly how to beat the Sharpe Ratio of an equally weighted asset allocation in our experiment.

[18] The results in this paper are based on different experimental sessions which might not allow for an aggregation of results from different sessions. What could have biased the results is that the Sharpe Ratio of the best participant was shown to all other participants in each round of an experiment. Theoretically the odds of having a lucky participant who starts with a very good allocation is significantly higher in experiments with 24 individuals than it is with only 5 groups so the feedbacks participants receive might differ. But there is one argument which invalidates this issue. We observed that in each group experiment at least one group was achieving very high Sharpe Ratios during each round of the experiment so the feedbacks were almost identical for all participants in every experimental session. Mann-Whitney test show that there are no significant differences between the best individuals and groups of each session. Furthermore, the number of observations in a single group session (5 groups) is too small to make any analysis so an aggregation is necessary to draw meaningful conclusions.

4.5.2 Performance comparison with stochastic groups

In this section we address hypotheses 1.2 and 1.3 which compare the quality of group decision-making relative to stochastic groups.[19] To judge the performance of groups in more depth we constructed two types of stochastic group samples mimicking potential group decision-making rules and we compared the Sharpe Ratios of those two stochastic group samples with our individual and group samples. One stochastic group sample served as a proxy for averaging the asset allocation decision of three individuals and was named "average stochastic group". An observation of the average stochastic group sample can be interpreted as a group decision in which the members of the group agreed to an even-handed compromise. To construct one observation in this sample we randomly selected the asset allocation of three individuals per round from our sample of 45 individuals and we calculated the average of those three allocations. The other stochastic group sample reflected a best-in-group decision and was called "best-in-group stochastic group". To construct an observation in this second sample we simply chose the best individual allocation from a group of three randomly chosen individuals per round, again from the sample of our 45 individuals. An observation in this sample served as a proxy for a group whose members were able to select their best individual allocation decision in each round. This follows the idea of Lorge and Solomon (1955) about the truth-wins-standard but in contrast to the original idea our approach did not require the optimal solution but only the best individual solution across all group members.

Both samples of stochastic groups contained 1000 observations and the random combinations of three individuals were unchanged over all rounds of the experiment. For example the stochastic group X consisted of individual 4, 15 and 33 during every round of the experiment.[20] To test hypotheses 1.2 and 1.3 we then compared the Sharpe Ratios of those two types of stochastic groups with the Sharpe Ratios of groups in each round.

---

[19] For an introduction into approaches that involve stochastic groups see Lorge, Irving , D. Fox, J. Davitz, and M. Brenner, 1958, A survey of studies contrasting the quality of group performance and individual performance, *Psychological Bulleting* 55, 337-372..

[20] There is one caveat in the method of constructing stochastic groups. The individuals that make up one observation in a stochastic group have received the feedback for their own personal asset allocation in each round of the experiment and not the feedback for the performance of the stochastic group. So it might be the case that individuals reacted differently if they received the stochastic group's feedback instead of the feedback for their individual asset allocation. On the other hand it is unclear in which way individuals had reacted. So a simple approach looks adequate to us because our stochastic groups only serve as a benchmark to judge the decision quality of individuals and groups.

Figure 4.1a shows that the average stochastic group sample as well as the best-in group stochastic sample outperformed our groups in almost all rounds of the experiment. In contrast figure 4.1b reveals that the median Sharpe Ratio in the group sample performed similarly to the median of the average stochastic group sample but below the best-in-group sample. This is evidence that large outliers in the experiments can be found more among the worst than the best groups and individuals. The differences between the average and the median Sharpe Ratios of stochastic groups were barely visible because those samples with 1000 observations followed a normal distribution. According to Mann-Whitney tests, the difference between the Sharpe Ratios of average stochastic groups and our real groups is only significant in 2 out of 15 rounds. So we cannot reject hypothesis 1.2 namely that groups perform similarly to average stochastic groups. As expected the average Sharpe Ratio of best-in-group stochastic groups was significantly higher than the average Sharpe Ratio of our 16 real groups in each but one round of the experiment at the 5% level so we cannot reject hypothesis 1.3.

In short, the sample of our real groups performed as well as the sample of average stochastic groups but clearly failed to reach the performance of best-in group stochastic groups. So group discussions did not seem to enhance the performance because our experiment showed that an average allocation of three randomly chosen individuals yielded no significantly different results.

In addition we observed that the sample of average stochastic groups realized significantly higher Sharpe Ratios than the individual sample in 9 out of 15 rounds at the 10% level and the average Z scores of all 15 rounds is -1.72.[21] This result occurred because averaging the Sharpe Ratios of individual allocations is not equal to calculating the Sharpe Ratio based on average asset allocations of 3 randomly chosen individuals. In our data roughly 70% of all observations showed the average Sharpe Ratio of 3 individuals below the Sharpe Ratio of the average asset allocation of those 3 individuals. The reason is that building the average of 3 individual allocations moderates extreme positions with very low Sharpe Ratios in our sample (for example the position of 100% in the riskiest asset A which yields a Sharpe Ratio of 0.42). So extreme allocations with low Sharpe Ratios occurred fewer times in the sample of average stochastic groups than

[21] The Sharpe Ratio of the best-in-group stochastic group sample is significantly higher than the Sharpe Ratio of the individual sample which is logical due to the construction method.

in the sample of individuals. A comparison between median and average Sharpe Ratios for the individual sample supports this observation especially in the first rounds of the experiment. The reason is that the average is lower than the median due to a few unsuccessful outliers among the individuals whose impact is moderated if they are combined with other individuals like in the sample of average stochastic groups.

4.5.3 Portfolio risk and asset allocation shifts

We now focus on the degree of risk in group and individual portfolios.

Figure 4-2: Volatility in the samples
Figure 4.2 plots the average portfolio volatility of individuals, groups and both types of stochastic groups per round. The sharp increase in the average volatility in the group sample in round 6 can be explained by two groups who both allocated a high percentage number to the most risky asset (A) in that round. As expected the portfolio volatility of individuals is always above the portfolio volatility of both types of stochastic groups. The reason is that non-perfectly correlated combinations of individual portfolios are always reducing the volatility of the resulting portfolio. The effect is decreasing over the rounds because the diversification effect across individuals is also decreasing as most of the individuals hold comparable allocations towards the end of the experiment.

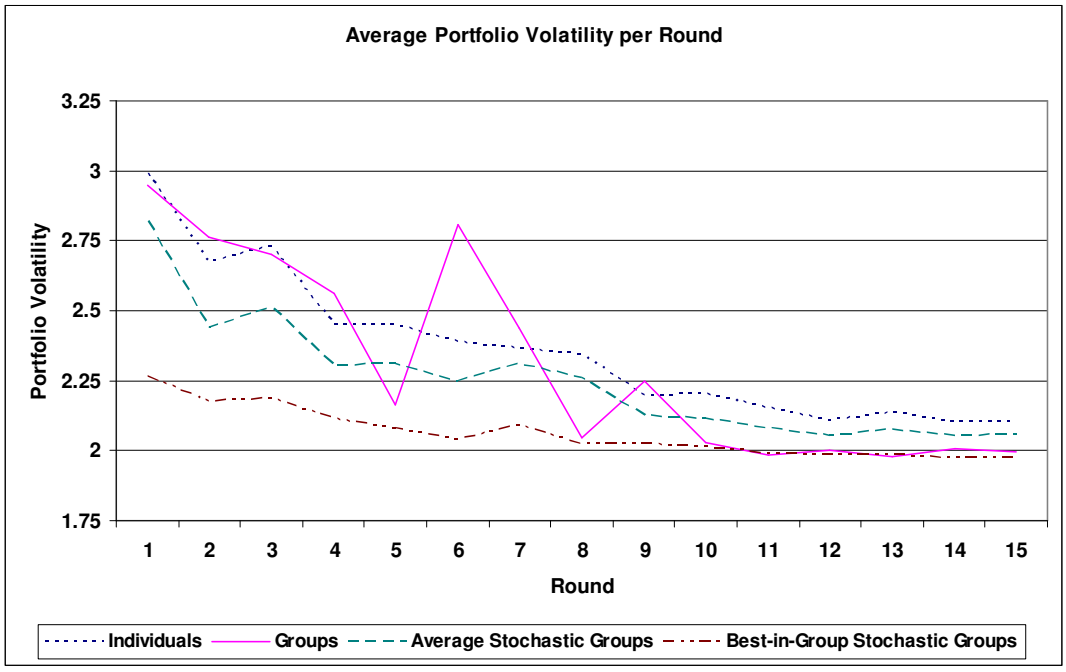
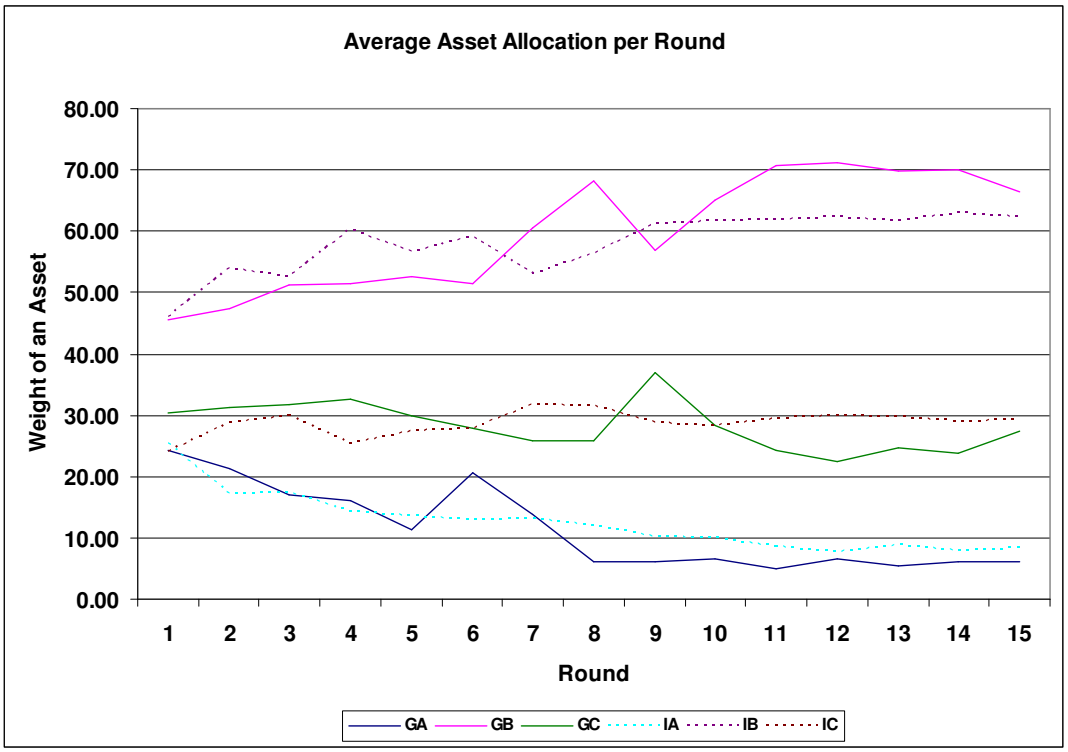


Figure 4.2 shows that the average portfolio volatility of groups was comparable to the one for individuals in most of the rounds. Mann-Whitney tests indicated that there were no significant differences between individuals' and groups' portfolio volatility at the 5%

level in all but two rounds of the experiment. The spike of the average group volatility in round 6 can be explained by two groups who both allocated a high percentage number to the most risky asset (A) in that round. So there is no evidence for a difference between individuals' or groups' preferences as far as portfolio risk is concerned and we cannot reject hypothesis 2. Figure 4.3 supports figure 4.2 in that there were no material differences between groups' and individuals' chosen asset allocations on average in each round. Both, groups and individuals, started with an overweight in the riskiest asset (A), an underweight in the least risky asset (B) and quite a good allocation to the mid risky asset (C).

Figure 4-3: Asset allocations

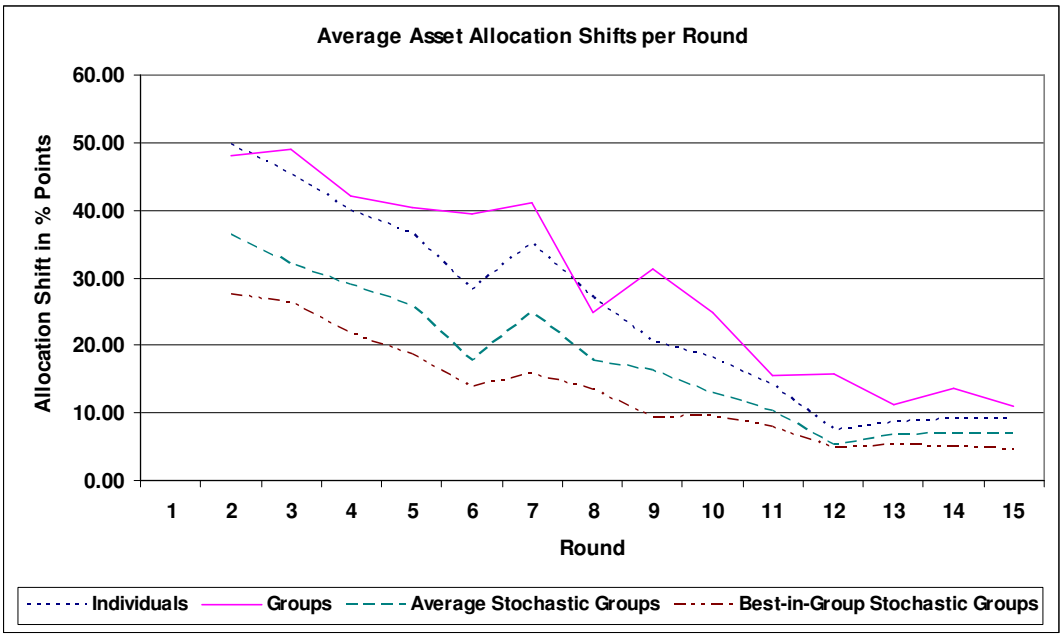
Figure 4.3 shows the average allocation of groups and individuals to the three risky assets (A, B and C) in each round of the experiment. The allocations in figure 4.3 have been normalized in order to make them comparable because the allocation to the risk-free asset (D) does not influence the Sharpe Ratio. So the allocations to the 3 risky assets have been inflated to a total of 100% while keeping the relative weighting of all 3 risky assets proportional. GA stands for groups' average allocation in asset A, IA for individuals' average allocation in asset A and so on.



From the choice of an allocation we now move on to the shifts of asset weightings in the asset allocation from one round to the next. There was no material difference between

the amounts of the shifts in the individual and the group sample as figure 4.4 points out. A shift represents the sum of all percentage points shifted from the allocation of the last round to the allocation in the next round. For example an allocation of A 30%, B 30% and C 40% in round 1 to an allocation of A 40%, B 30% and C 30% in round 2 represents a shift of 20%. Groups shifted a higher amount of percentage points than individuals in all but two rounds but the effect was insignificant in every round and therefore we cannot reject hypothesis 3. On average both individuals and groups shifted roughly 50% points in the first round and then almost linearly decreased this rate to roughly 10% points in the last round as can be seen in figure 4.4.

Figure 4-4: Shifts in the asset allocations
The figure 4.4 shows the median shifts in the asset allocation expressed as percentage points shifted from one asset to another per round for individuals, groups and both sorts of stochastic groups.



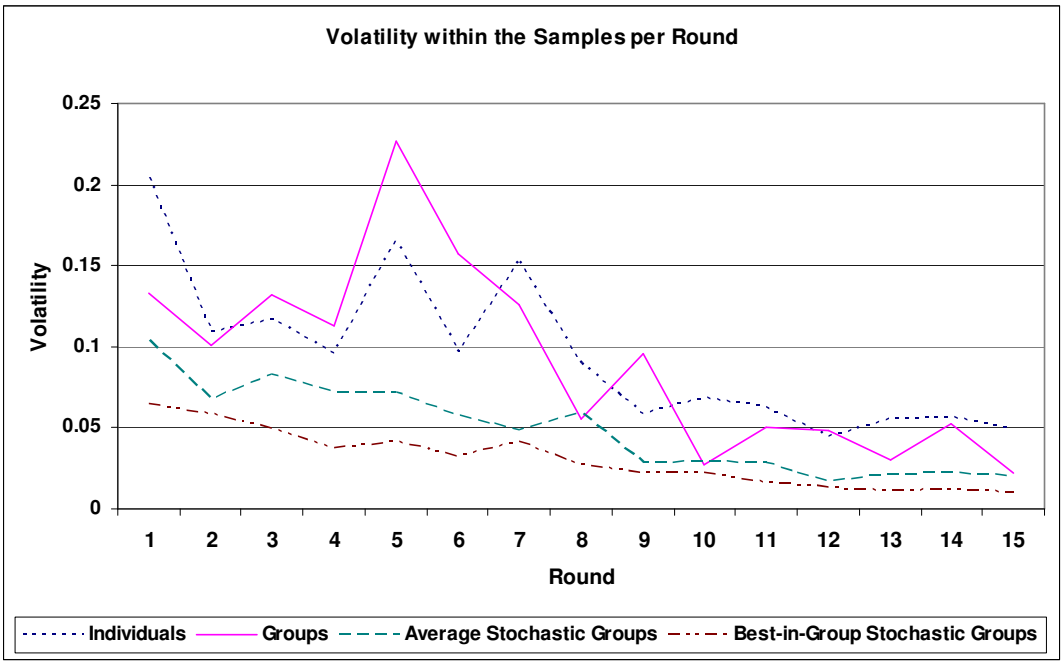
The results in this section led to the conclusion that groups do not choose less risky asset allocations and share the same level of flexibility with respect to new allocations.

4.5.4 Comparison of the performance volatility of the samples

So far we have only considered averages and medians in all of our samples. Now we focus on differences within the samples. To analyze the homogeneity of individuals and groups we compared the volatilities of the Sharpe Ratios in the different samples over

all rounds of the experiment. Figure 4.5 shows that for all samples the volatility decreased during the course of the whole experiment indicating that the participants' in each sample agreed more and more about the optimal allocation in the experiment from one round to the next.

Figure 4-5: Volatility of Sharpe Ratios
Figure 4.5 shows the volatilities of the Sharpe Ratios in the individual, the group and both of the stochastic group samples.

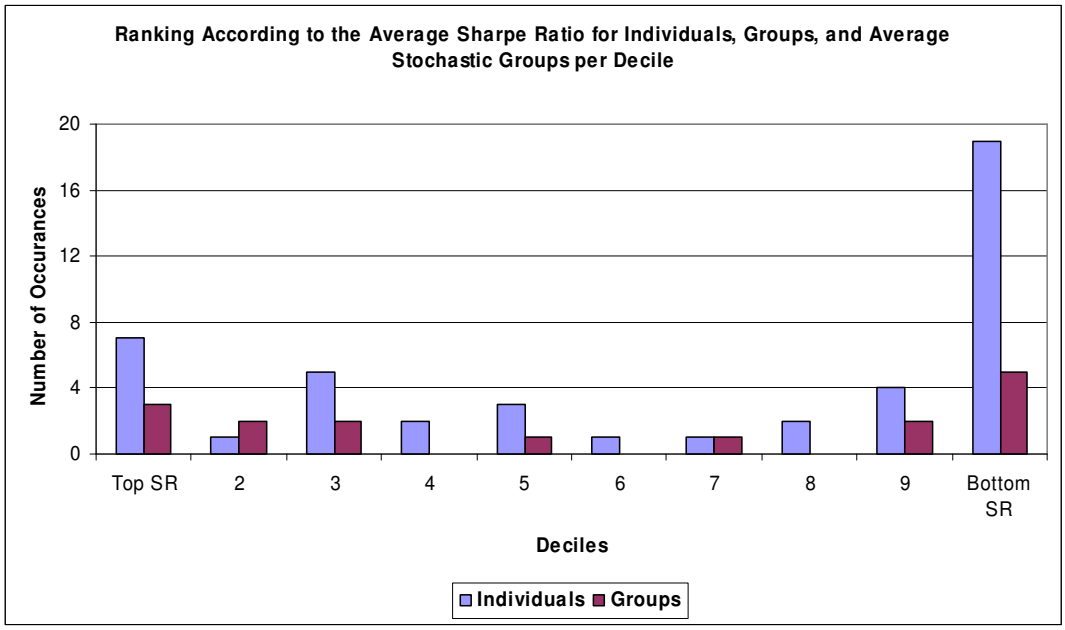


We found no significant difference between the sample volatility of individuals and groups in a Mann-Whitney test. But the difference between those two samples and the two sorts of stochastic group samples is significant at the 5% level. Therefore, the samples of individuals and groups showed a higher volatility than the samples of both types of stochastic groups over the whole experiment. A high volatility across the Sharpe Ratios reflects large discrepancies across the asset weightings and we will address potential reasons for those discrepancies in the next section.

To analyze the discrepancies in the performance in more detail one comprehensive sample that consisted of the individuals, the groups and average stochastic groups was constructed. We then split this comprehensive sample into deciles according to the average Sharpe Ratio over all 15 rounds of the experiment and we analyzed the structure of each decile (figure 4.6).

Figure 4-6: Ranking of Sharpe Ratios

The individuals', groups' and average stochastic groups' samples have been aggregated into one single sample and then deciles according to the average Sharpe Ratios (SR) over all 15 rounds have been calculated. In each decile there are 106 observations. The figure 4.6 shows the number groups and individuals in each of those deciles from the top Sharpe Ratios to the bottom Sharpe Ratios.



It can be observed that the average Sharpe Ratios over the whole experiment of most individuals as well as most groups were located either in the top or in the bottom decile. Given the structuring approach of average stochastic groups this is consistent with our previous arguments that an aggregation of individual allocations will lead to a moderation of extreme positions. It is interesting to note also, that 12 out of 16 groups were located either in the top two deciles or the bottom two deciles and only 4 groups fell into the 6 deciles in the middle. The 7 groups in the lowest 2 deciles did not moderate extreme individual positions as much as our sample of average stochastic groups. This suggests the presence of a group polarization effect because the unsuccessful groups in our sample did not build asset allocations that led to Sharpe Ratios in the middle of the sample distribution but made rather bad allocations compared to average stochastic groups.

A sensitivity analysis reveals that the sample distribution of the average Sharpe Ratios from the first 5 rounds, the rounds 6-10 and the last 5 rounds were not different and the distribution of individuals and groups across the deciles did not change materially. It

follows that the observation of large performance discrepancies is stable over the whole experiment as well as over those 5 round periods.

4.5.5 Decision-making processes of the groups

From the section before we know that the volatility across groups was high (figure 4.5) and that most of the groups either performed very well or very badly (figure 4.6). In this section we test hypothesis 4 which offers an explanation for those results. The basic idea is that the level of information exchange across group members is positively related to the performance of a group. The recordings of the group discussions reveal that the intensity and the frequency of the information exchange are different across the groups in our sample. We decided to judge the relationship between the level of information exchange and the performance with four factors which are listed in table 4.2. Those are quantifiable factors in order to measure the true information exchange level as objectively as possible. The scores of each factor then allow a crude assessment of the level of information exchange across group members. We only rely on recordings between the start of the experiment and round 9 because the differences in the Sharpe Ratios from round 10 onwards are small because most of the groups have found a pretty good asset allocation after those first 9 rounds.

Table 4-2: Factors in the decision making processes of groups

Table 4.2 contains the four factors of the group decision-making process for each group (Focus Time, Ratio, Suggestions and Arguments) as well as the spare time used to calculated the factor "ratio". The corresponding scores and the aggregated Information Exchange Scores (IE Scores) are listed for each group except for group 5 because the discussion recording of this group is not available due to technical issues.

Table 4.2

Group	Focus Time in Minutes	Spare Time in Minutes	Ratio Focus Time / Spare Time	Number of Suggestions	Number of Arguments	Information Exchange Score
1	12	16	0.8	15	5	low
2	27	1	27.0	39	26	high
3	12	16	0.8	24	6	low
4	28	0	28.0	47	28	high
5	9	9	9.0	9	9	na
6	22	0	22.0	32	13	medium
7	16	6	2.7	33	15	medium
8	20	3	6.7	38	21	high
9	15	7	2.1	16	5	low
10	17	5	3.4	24	15	low
11	20	2	10.0	58	33	high
12	19	3	6.3	40	28	medium
13	23	0	23.0	43	34	high
14	16	7	2.3	27	13	low
15	19	4	4.8	32	13	medium
16	21	2	10.5	48	19	high

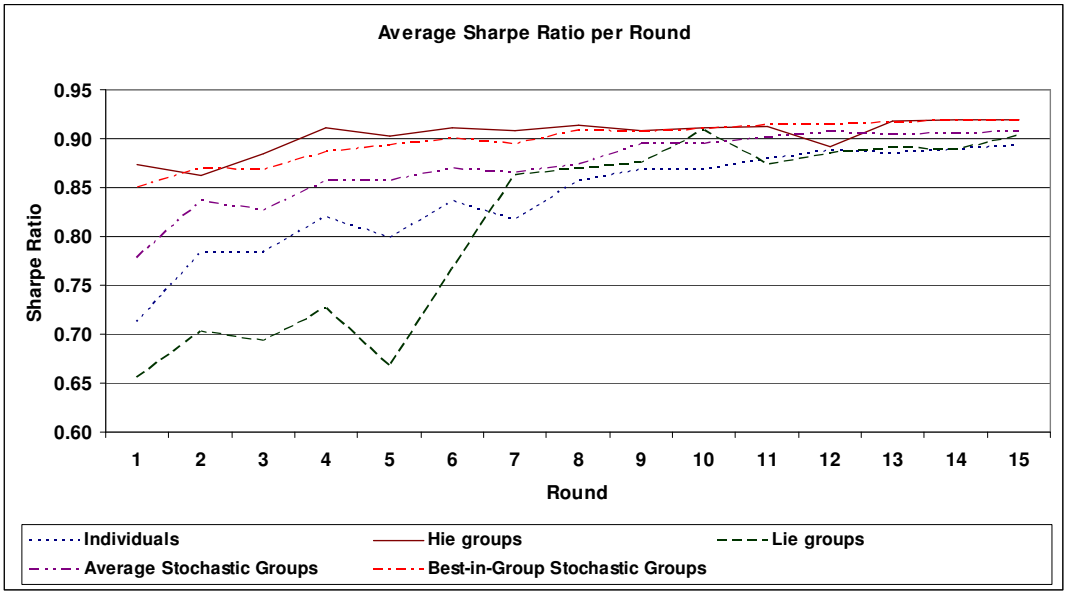
The first factor is called "focus time" and represents the minutes each group was focusing on the experiment by discussing various issues about how to solve the asset allocation problem. Factor two is named "ratio" and reflects the ratio of focus time and spare time where spare time is defined as the time the group discussed other topics than the experiment. The reason to include this factor is that not every experimental session was equally long and this factor serves as a measure to judge a group's focus on the experiment independently from the absolute level of focus time. The third factor "suggestions" measures how many suggestions each group has brought up internally by adding up every individual suggestion of every group member. A suggestion is defined as a statement of a group member about how much weight a certain asset should receive in the asset allocation. Factor four is named "arguments" and counts the number of arguments that have been brought up by the group members during the discussion. An argument is defined as a statement that supports a group member's suggestion in that it

either points out an advantage or demonstrates quality in order to convince the other group members about the suggestion.

Then the results of each factor were aggregated into a single measure called "information exchange score" (last column in table 4.2). Every group that scored above the median in all four factors was categorized as a high-information-exchange-group (HIE-group). In total 6 groups are categorized as HIE groups. Groups that scored below the median in all four factors were referred to as low-information-exchange-groups (LIE-group) and we found 5 such groups in our sample. 4 groups are not categorized because they performed above the median in some but not all four factors so we cannot judge if they belong to the HIE or the LIE group. One group was excluded because the recording was not available due to technical issues. Finally, the Sharpe Ratios of HIE-groups and LIE-groups were compared for every round of the experiment and Mann-Whitney tests served to analyze the significance of the differences.[22]

Figure 4-7: Sharpe Ratios of groups

Figure 4.7 contains the average Sharpe Ratios of individuals, high-information-exchange groups (HIE-groups), low-information-exchange groups (LIE-groups), average stochastic groups and best-in-group stochastic groups for every round in the experiment. The sharp fall of the average Sharpe Ratio in the LIE groups in round 5 was because of a 100% allocation to the riskfree asset of one group which had a strong negative impact. The curve for the LIE groups would be much smoother if we exclude this particular group in round 5.



[22] We acknowledge that the correlations across our four factors are pretty high with a range from 0.557 to 0.923.

Figure 4.7 shows that the Sharpe Ratios of HIE-groups were larger than the Sharpe Ratios for LIE-groups in every round of the experiment. The difference is significant according to Mann-Whitney tests at the 10% level in 10 rounds. HIE groups also beat individuals at the 10% significance level in 12 out of 15 rounds. Furthermore, HIE-groups significantly outperformed average stochastic groups in 12 out of 15 rounds at the 10% level and marginally beat the performance of the best-in-group stochastic groups in almost every round. In contrast the average Sharpe Ratio for LIE-groups was even lower than for individuals in the first 6 rounds but not on a significant level except for 2 rounds. The sharp fall of the average Sharpe Ratio in the LIE groups in round 5 was because of a 100% allocation to the riskfree asset of one group which had a strong negative impact. If we exclude this observation the performance of LIE groups in figure 4.7 would be much smoother but the disappointing results of LIE groups do not change materially.

It is not surprising that the largest differences between HIE and LIE groups occurred in the first rounds of the experiment because the information about the best performer in each round helped every participant to judge his performance round by round. While HIE groups identified efficient allocations pretty early in the experiment due the high information exchange, LIE groups needed more time to evaluate their performance.

In a sensitivity analysis we structured the HIE and LIE groups slightly different by adding the four uncategorized groups to either HIE or LIE groups. The results do not change materially as the sample of HIE groups always outperformed the sample of LIE groups. We conclude that in line with the thinking of Stasser and Titus (1985) a high level of information exchange in a group leads to a significantly better performance.

4.5.6 Decision-making time

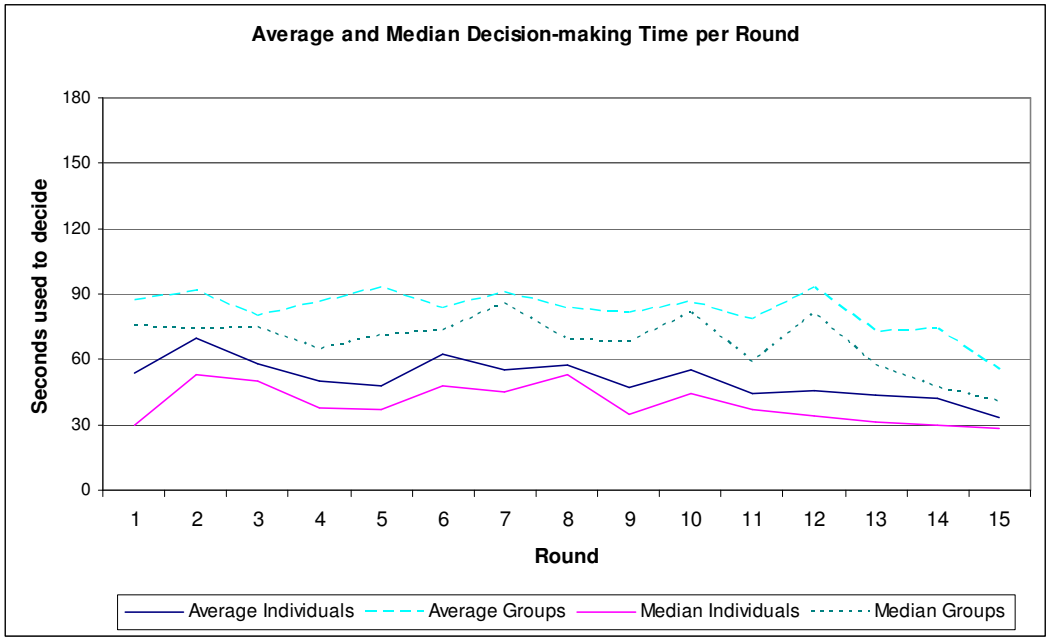
From the section above we know that the factors “focus time” and “ratio” are positively related to a better performance within the sample of groups. A comparison of the time used to enter an asset allocation decision on the computer terminal between HIE-groups and LIE-groups reveals that HIE-groups used more time in every round of the experiment and Mann-Whitney tests confirm that the difference is significant in 10 rounds at the 10% level. Maybe groups who exchange more information perform better because a

higher level of information exchange is linked to more time to analyze the allocation task in our experiment.

Individuals had no chance to exchange any information so it is not surprising that they needed much less time to enter their asset allocation. On average individuals used 51 seconds to decide whereas groups used 82 seconds per round. Mann-Whitney tests prove that the difference between groups' and individuals' was significant at the 5% level in 14 out of 15 rounds. Figure 4.8 shows the average and the median decision time of individuals and groups per round. The time used to define the allocation decreased for both groups and individuals during the experiment but the difference is stable.

Figure 4-8: Decision-making time

Figure 4.8 shows the average and the median time used to decide about the allocation for groups and for individuals per round.



4.5.7 Performance beliefs of the participants

Finally, we focused on the participants' ex post opinions about the performance of groups versus individuals in our experiment. After the completion of their experimental session in total 45 students expressed the view in questionnaire II that groups will achieve higher Sharpe Ratios in the asset allocation task and only 18 students believed individuals will outperform groups. The other participants expressed no opinion. This

observation was much stronger across group members as only 4 out of 34 group members believed individuals would outperform in the experiment whereas 14 out of 29 of the individuals shared that belief. It seems as if the participants in the group experiment put value on the fact that they could solve the experiment as a group. But the achieved Sharpe Ratios in the experiment only confirm the group member's impression in HIE-groups.

4.6 Discussion and conclusion

Groups on average outperform individuals in an asset allocation task with intellective features but the difference is not significant. However, an analysis of the group sample average disregards the heterogeneity of the groups' performances. Groups with a high level of information exchange performed as well as best-in-group stochastic groups which by construction selected the best individual allocation of 3 randomly chosen individuals. Therefore, groups marginally beat even a special version of the truth-wins-standard. On the other hand groups with a low level of information exchange underperformed all other samples and the difference was significant in the first rounds of the experiment.

Our results support the existing literature by providing evidence that the decision-making process, in our case addressed via the information exchange level, significantly influences the success of groups. In line with the thesis of Janis (1982) this paper shows that successful groups evaluate available information and ideas of all group members before choosing an asset allocation. In contrast unsuccessful groups typically opt for the first suggestion by a group member and therefore do not make use of the possibility to pool their personal information. Our results also emphasized the hidden profile issue by Stasser and Titus (1985) from another angle as we demonstrated that a lack of information exchange is related to an underperformance of groups. It is beyond the scope of this paper to analyze reasons for the differences across the information exchange level in the groups in depth, but further research is needed to figure out how certain factors influence group members' willingness to exchange information.

One potential answer lies in the group polarization effect of Moscovici and Zavalloni (1969). The worst performing groups achieved Sharpe Ratios significantly below those of stochastic groups mimicking a decision-making process based on even-handed com-

promises between the group members. It seems that the absence of information exchanges within the group does not seem to moderate individual opinions, i.e. asset weightings in the asset allocation, but rather produces extreme asset allocations. So groups who do not carefully discuss potential solutions but agree to the first suggestion of one group member might implement more extreme asset weightings than the aggregated opinion of all group members.

After the experiment a large part of all participants shared the view that groups should outperform individuals. It might be the case that people simply prefer to take decisions in a group regardless of the decision-making process and the achieved results. One reason could be the delegation of responsibility from an individual to the group where potential harm due to a bad decision is shared in a group and therefore no single individual is blamed completely. Also it is probably less disappointing to be wrong as part of a group than being wrong individually. However, more research is needed to analyze the confidence of individuals to better solve problems in a group.

Indeed in reality many asset managers organize their decision-making process of fund management in groups as Baer, Kempf and Ruenzi (2005) report. However, further research is needed to analyze people's preferences in the domain of financial decision-making in general. With respect to group decision-making in practice we point out that group decisions do not add value per se unless all the group members' opinions are communicated and discussed. So whenever group decisions need to be taken it is important to create an environment which motivates people to exchange information.

Another meaningful extension to our experiment would be to introduce a judgmental element by letting the participants trade the assets they use for the asset allocation task during the experiment. In such a task the participants not only have to find the optimal solution for the asset allocation problem but do also have to take the preferences and actions of other participants into account. An interaction of participants reflects reality much better because the prices and the returns of assets on financial markets are not stable as in our experiment but heavily dependent on the investors' supply and demand situation. Adding an interactive element onto our experimental design might add more insight into the decision-making process, the rationality, and the potential success of individuals and groups on financial markets.

4.7 Literature

Baer, M., Kempf, A., and Ruenzi, S. "Team Management and Mutual Funds." *CFR Working Paper*, No. 05-10, 2005.

Barber, B., and Odean, T. "Too Many Cooks Spoil the Profits: Investment Club Performance." *Financial Analysts' Journal,* (2000), pp. 17-25.

Blinder, A., and Morgan, R. "Are Two Heads Better Than One? Monetary Policy by Committee." *Journal of Money, Credit, and Banking,* 37, (2005), pp. 798-811.

Bornstein, G., Kugler, T., and Ziegelmeyer, A. "Individual and group decisions in the centipede game: Are groups more "rational" players?" *Journal of Experimental Social Psychology,* 40, (2004), pp. 599-605.

Bossaerts, P., and Plott, C. "The CAPM in thin experimental financial markets." *Journal of Economic Dynamics and Control,* 26,, (2002), pp. 1093-1112(1020).

Cason, T., and Mui, V.-L. "A laboratory study of group polarization in the team dictator game." *Economic Journal,* 107, (1997), pp. 1465-1483.

Cooper, D., and Kagel, J. "Are Two Heads Better than One? Team versus Individual Play in Signaling Games." *American Economic Review,* 95, (2005), pp. 477-509.

Davis, J. "Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950-1990." *Organizational Behavior & Human Decision Processes,* 52, (1992), pp. 3-38.

Fischbacher, U. "z-tree. Toolbox for Readymade Economic Experiments." *IEW Working Paper, University of Zurich*, 21, 1999.

Greiner, B. "The Online Recruiting System ORSEE 2.0. A Guide for the Organization of Experiments in Economics." *Working Paper Series in Economics, Department of Economics, University of Cologne*, 10, 2004.

Janis, I. "Groupthink: Psychological Studies of Policy Decisions and Fiascoes." Houghton Mifflin, Boston, 1982.

Kerr, N., Kramer, G., and MacCoun, R. "Bias in Judgement: Comparing Individuals and Groups." *Psychological Review,* 103, (1996), pp. 687-719.

Kocher, M., and Sutter, M. "The Decision Maker Matters: Individual versus Group Behaviour in Experimental Beauty-Contest Games." *Economic Journal,* 115, (2005), pp. 200-223.

Laughlin, P., Bonner, R., and Altermatt, T. "Collective versus Individual Induction with single versus multiple hypothesis." *Jounal of Personality and Social Psychology,* 75, (1998), pp. 1481-1489.

Lombardelli, C., Proudman, J., and Talbot, J. "Committees versus Individuals: An Experimental Analysis of Monetary Policy Decision Making." *International Journal of Central Banking,* 1, (2005), pp. 181-207.

Lorge, I., Fox, D., Davitz, J., and Brenner, M. "A survey of studies contrasting the quality of group performance and individual performance." *Psychological Bulleting,* 55, (1958), pp. 337-372.

Lorge, I., and Solomon, H. "Two models of group behavior in the solution of eureka-type problems." *Psychometrica,* 20, (1955), pp. 139-148.

Moscovici, S. "Social influence and conformity." in Lindzey, G., and Aronson, E., eds.: Handbook of Social Psychology. 1985.

Moscovici, S., and Zavalloni, M. "The group as a polarizer of attitudes." *Journal of Personality and Social Psychology,* 12, (1969), pp. 125-135.

Rockenbach, B., Sadrieh, A., and Mathauschek, B. "Teams take the better risks." *Journal of Economic Behavior and Organization,* Forthcoming, (2006), pp.

Sah, R. K., and Stiglitz, J. "The Architecture of Economic Systems: Hierarchies and Polyarchies." *American Economic Review,* 76, (1986), pp. 716-727.

Schulz-Hardt, S., and Greitmeyer, T. "Preference-Consistent Evaluation of Information in the Hidden Profile Paradigm: Beyond Group-Level Explanations for the Dominance of Shared Information in Group Decisions." *Journal of Personality and Social Psychology,* 84, (2003), pp. 322-339.

Shaw, M. "A Comparision of Individuals and Small Groups in the Rational Solution of Complex Problems." *American Journal of Psychology,* 44, (1932), pp. 491-504.

Stasser, G., and Titus, W. "Pooling of unshared information in group decision making: Biased information sampling during discussions." *Journal of Personality and Social Psychology,* 48, (1985), pp. 1467-1478.

Stoner, J. "A comparison of individuals and group decisions involving risks." *Thesis (M.S.)-Massachusetts Institute of Technology, School of Industrial Management*, 1961.

Yaniv, I., and Bornstein, G. "Individual and group behaviour in the ultimatum game: Are groups more rational players?" *Experimental Economics,* 1, (1998), pp. 101-108.

4.8 Appendix I

The instructions provided in this section are translated from the original German instructions.

Instructions to the experiment

Welcome to our experiment! In this experiment you will be asked to take several decisions on a computer terminal. Your payoff depends on your decisions and on the decisions of the other participants. So please read the following instructions carefully.

In the experiment certain technical expressions are used that might not be familiar to you. Those expressions will be introduced on the next pages of this instruction. In case you have any questions please raise your hand and one of the instructors will come to your place and answer your questions.

The participants in this experiment are students with different backgrounds from the University of Zurich and from the ETZH. All participants receive exactly the same information like you. All participants can keep and use the instructions during the whole experiment.

The experiment starts as soon as every participant has fully understood the course of the experiment and has correctly completed the questionnaire I.

Please respect that communication with the other participants is not allowed during the experiment. Mobile phones must be switched off. Please operate only those functions on the computer which are needed in the course of the experiment. Communication with other participants or bugger with the computer leads to an exclusion of the experiment.

In case you have any question the instructors are ready to help.

Goal of the experiment

The goal of the experiment is for all participants to combine 4 different assets (3 risky assets A, B, C and 1 riskfree asset D) in such a way to achieve as much return as possible with as small risk as possible over all rounds of the experiment. The experiment consists of 15 equal rounds in which all participants can change the allocation across the 4 different assets in order to increase their risk-return-ratio.

All relevant technical expressions of the experiment will be explained in the next section.

Explanations

Asset: An asset is an investment that increases (or decreases) in value. In a riskfree asset (like for example a bank account) the return is fixed. In a risky asset like for example a stock the return is variable.

Return: The return is defined as the increase or decrease in value of an asset expressed in percentage points. An asset with a return of 5% increases its value by 5% and an asset with a return of -1% reduces its value by 1%.

Scenarios: The return of an asset in the experiment depends on the occurance of 4 different scenarios (I, II, III or IV). The probability of an occurrance of a scenario is equal (i.e. 25%) for all scenarios. The following table provides an overview about the returns of the different assets A, B, C and D in the different scenarios.

The table is at your disposal on the computer screen during the whole experiment.

Scenario	Probability of occurance	Return Asset A	Return Asset B	Return Asset C	Return Asset D
I	0.25	15%	4%	0%	1%
II	0.25	-2%	5%	5%	1%
III	0.25	8%	1%	9%	1%
IV	0.25	-4%	0%	-1%	1%

Example 1: In scenario I asset A returns 15%, asset B returns 4%, asset C returns 0% and asset D returns 1%.

Example 2: In Scenario II a combination of 50% in asset A, 50% in asset B and nothing in assets C and D returns 1.5% (0.5 * (-2)% + 0.5 * 5% = 1.5%).

Mean Return: The mean return is the return that is achieved with an asset or a combination of assets in the mean across different scenarios.

Example 3: The mean return of asset B across all 4 scenarios is

0.25 * 4% + 0.25 * 5% + 0.25 * 1% + 0.25 * 0% = 2.5%

Example 4: The mean return of a combination of assets with a weight of 50% in B, 50% in D and nothing in A and C is

0.25 * (0.5 * 4% + 0.5 * 1%) + 0.25 * (0.5 * 5% + 0.5 * 1%) +
0.25 * (0.5 * 1% + 0.5 * 1%) + 0.25 * (0.5 * 0% + 0.5 * 1%) = 1.75%

Excess return: The mean return of an asset or a combination of assets minus the riskfree return of the asset D yields the excess return:

Excess return = Mean return of an asset – Riskfree return

Example 5: The excess return of asset B is 2.5% - 1% = 1.5%.

Example 6: The excess return of a combination of assets with a weight of 50% in B, 50% in D and nothing in A and C is 1.75% - 1% = 0.75%.

Risk: Risk is defined as the variability of the returns of an asset or a combination of assets in different scenarios. The larger the variability is the higher is the risk. The risk is calculated by taking the square root of the sum of the squared differences between scenario returns and the mean return of an asset. When we call the scenario returns of an asset R_I, R_{II}, R_{III}, R_{IV} in the different scenarios and M the mean return of this asset then the formula for the risk is:

Example 7: If we assume, $R_I = R_{II}$=5% and $R_{III}=R_{IV}$=10%. Then the mean return is

M=0.25 * 5% + 0.25 * 5% + 0.25 * 10% + 0.25 * 10%= 7.5% and the risk is

$$Risk=\sqrt{0.25(5\%-7.5\%)^2+0.25(5\%-7.5\%)^2+0.25(10\%-7.5\%)^2+0.25(10\%-7.5\%)^2}$$

Risk-return-ratio: The risk-return-ratio is defined as the ratio between the excess return of an asset and the risk of an asset or a combination of assets:

Risk-return-ratio = Excess return / Risk

The risk-return-ratio increases if the excess return increases or if the risk decreases.

Special case: If the risk is 0 then the risk-return-ratio is also 0.

Example 8: An asset with a mean return of 5% and a risk of 8% offers a risk-return-ratio of (5% - 1%) / 8% = 0.5.

Payoffs

Your payoff is linked to your performance in the experiment. The relevant measure is the risk-return-ratio. Based on your achieved risk-return-ratio in each round the mean risk-return-ratio for the whole experiment is calculated. During the experiment you see permanently the level of your average risk-return-ratio across all rounds played so far.

Your average risk-return-ratio will be multiplied with 30. This will yield the payoff in CHF that you receive at the end of the experiment.

The participant with the highest average risk-return-ratio receives an additional bonus of 20 CHF. For the participants ranked 2., 3., and 4. with respect to the risk-return-ratio a bonus of 10 CHF will be paid. In addition to the performance based payoff every participant receives a show-up bonus of 10 CHF.

The payoffs are paid immediately after the experiment by the instructors.

Experimental design

The experiment consists of 15 rounds and one trial round.

Before the trial round questionnaire I must be correctly completed. Hand in your questionnaire I to the instructors as soon as you have it completed.

Every participant has to choose in each round how he wants to weight the 4 different assets A, B, C and D to achieve the highest possible risk-return-ratio. Every participant can allocate 0% to 100% to each asset (enter the weight without decimal places into the computer). The sum of all asset weights must add up to 100%.

As soon as all participants have entered their combination of assets the computer will calculate the mean return, the risk and the risk-return-ratio of each participant.

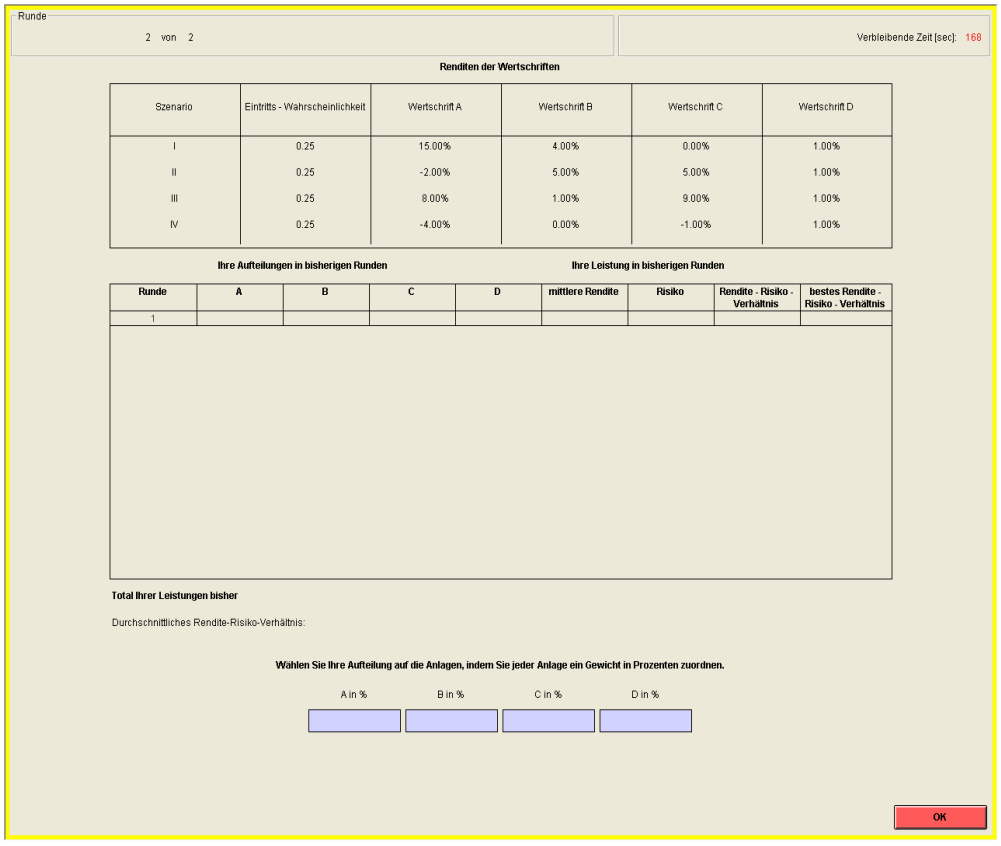
Every participant will be informed about his own numbers on the computer screen. In addition every participant gets informed about the highest risk-return-ratio that was achieved by a participant in the round. With this a round is completed. At the beginning of the next round each participant can chose a new weight for each of the 4 assets. This process will be repeated until 15 rounds have been played.

Course of the experiment

Please show your answers to an instructor as soon as you have completed questionnaire I. The trial round starts as soon as every participant correctly answered all questions in questionnaire I.

Trial round

In the trial round you can make yourself familiar with the computer and the different actions you need to take during the experiment. On the screen top left shows the round that is played and top right how much time you have left to enter your asset weights.



The upper table shows the return of each asset A, B, C, D in the different scenarios of the experiment.

The table in the middle shows your weightings of each asset in the previous rounds and the corresponding return, risk and risk-return-ratio you have achieved with your allocation. It also shows the highest risk-return-ratio of a participant in a round of the experiment. Each completed round will be listed in this table.

Your overall performance is the basis for your payoff at the end of the experiment. The average risk-return-ratio is the relevant measure and it is shown below the middle table of the screen.

In each round you see 4 small boxes on the bottom of the screen. Please enter the weightings of each asset for the next round in those boxes. Check your weights and make sure that they add up to 100%. Then please confirm by clicking ok. You have 3 minutes to enter your asset allocation in each round. If you have not entered an allocation for each asset within 3 minutes 100% will be booked into the riskfree asset D.

As soon as all participants completed the trial round the experiment starts with round 1.

Rounds

The course of each round is similar to the trial round and the screens also look identical. As soon as all the participants enter their asset allocation you will receive the information about your performance in the round.

End of the experiment

After round 15 the experiment is over. The instructors will inform you accordingly and hand out the payoffs.

4.9 Appendix II

The questionnaires provided in this section are translated from the original German questionnaires. It is important to note that technical expressions like return, risk or volatility have been introduced and defined in the instructions (see appendix I). In addition every participant had the chance to ask questions before completing the questionnaire I.

Questionnaire I – Questions before the start of the experiment

1. In 4 scenarios a certain combination of assets A, B, C and D yielded the following returns:

   3% -1% 6% 4%

   What is the mean return of this combination of assets?

3%	5%	6%	1%
O	O	O	O

2. What is the excess return of this combination of assets with a riskfree rate of 1%?

4%	3%	2%	0%
O	O	O	O

3. Which of the following two assets is riskier?

Asset X with the following returns in the 4 scenarios of the experiment	Asset Y with the following returns in the 4 scenarios of the experiment
12%, -6%, 9% and -1%	3%, 0%, 2% and -1%
O	O

4. Which of the two combinations of assets offers the higher risk-return-ratio?

Combination X with an excess return of 5 % and a risk of 10%	Combination Y with an excess return of 4% and a risk 4%
O	O

5. What is the payoff of a participant if he achieved an average risk-return-ratio of 0.5 in the experiment (without any bonus and without the show up fee)?

20 CHF	15 CHF	12 CHF	10 CHF
O	O	O	O

6. Did you understand the goal and the rules of the experiment?

   O Yes O No

Questionnaire II – Questions after the experiment

1. Age:.....

2. Gender: O Female O Male

3. What do you study?

Natural Science (ETHZ)	Anything but natural science (ETHZ)	Economics (UniZH)	Anything but economics (UniZH)
O	O	O	O

4. Who do you think will achieve higher risk-return-ratios in the experiment, groups or individuals?

Groups	Individuals	Similar	Equal
O	O	O	O

5. Would you have preferred to participate as an individual (as a group member) in the experiment?

O Yes O No O Does not matter

6. How do you judge your risk-return-ratio relative to the other participants in the experiment?

Very low	Low	Rather low	Average	Rather high	High	Very high
O	O	O	O	O	O	O

7. How did the risk-return-ratio of your asset allocations evolved during the experiment?

Surely increased	Increased	Rather increased	Constant	Rather decreased	Decreased	Surely decreased
O	O	O	O	O	O	O

8. Do you have a private securities account?

O Yes O No

9. How often do you read financial news or financial articles?

Daily	Almost daily	Weekly	Monthly	Quarterly	Yearly	Never
O	O	O	O	O	O	O

10. How would you judge your interest in financial topics?

Very low	Low	Rather low	Average	Rather high	High	Very high
O	O	O	O	O	O	O

5. Concluding Remarks

With these concluding remarks I would like to point out again the main contributions of my PhD Dissertation to research and to provide some advice for practitioners. In addition I also address some limits of my results and provide ideas for future research.

The main contribution of the first two essays to academic research is to empirically show that overconfidence is not only a phenomenon across private investors and mutual fund managers but also across a sample of Swiss pension fund managers. However, I have demonstrated that the degree of overconfidence in the domain of financial markets is different across individuals and so a calculation of a sample average disregards the heterogeneity of the overconfidence phenomenon. Personal characteristics like education, experience and age influence a person's degree of overconfidence. Nevertheless, overconfidence generally serves as an explanation for the popularity of active management at Swiss pension plans as many managers seem to be prone to the better-than-average-effect. Given the mediocre track record of active management at Swiss pension plans on average such expectations seem to be very optimistic. The most important recommendation to practitioners therefore is to get aware of overconfidence effects like miscalibration and the better-than-average-effect and to take those into account when estimating returns on financial markets and when analyzing the capabilities of asset managers.

The third essay contributes to the existing literature by proving that groups do not outperform individuals per se but that the information exchange across group members is a critical factor for success. Therefore, the recommendation to practice is to create an atmosphere in a group decision-making process which motivates all group members to exchange all available personal information.

However, there are a couple of arguments and ideas that need to be addressed on a very general level not only to put the different results of my research efforts into perspective but also to outline further potential areas for research.

We know that overconfidence can lead individuals to formulate biased views regarding the performance of assets on financial markets and the performance of their own asset

managers. But bear in mind that there is no way to link the specific performance of one pension plan to a participant's specific degree of overconfidence. In other words, we cannot prove any relationship between subjective degrees of overconfidence and the corresponding performance on financial markets because we only observe aggregated data. The conclusions of the first two essays therefore refer to the sample as a whole but we can not rule out that some participants correctly believe themselves to be above average and to be able to forecast future returns very precisely. Exactly because of this heterogeneity regarding individual levels of overconfidence further research should focus on empirical evidence about the relationship between subjective overconfidence and subjective performance on financial markets.

Furthermore, the linear regression model in the first essay only captures roughly 20% of the variability of the participants' confidence intervals. So it would be interesting to search for additional factors which might influence a person's subjective degree of overconfidence and to analyze the relationship between an investor's degree of overconfidence and his/her performance on financial markets in more depth.

The third essay presents evidence that the level of information exchange across group members is significantly related to the performance of the group in my market experiment. However, the results rely on the recordings of the group discussions and it appears highly plausible that other factors which have not been captured in my approach influence the performance of groups. More detailed observations of group decision-making processes are needed to shed some light on these driving forces. Also the experimental design in the third essay is a purely intellective task but in reality the asset allocation on real financial markets also involves judgmental elements. Another meaningful extension to the experiment would be to introduce a judgmental element by letting the participants trade the assets they use for the asset allocation task during the experiment. An interaction of participants reflects reality much better because the prices and the returns of assets on financial markets are not stable as in our experiment but heavily dependent on the investors' supply and demand situation.

On the one hand I hope to have motivated other PhD candidates to continue the research efforts in the topics I have addressed in my collection of three essays and on the other hand I hope to provide some practical input about how to take better financial decisions.

Curriculum Vitae

Personal Details

Name / Surname	Gort Christoph
Place, Date of Birth	Berne, January 18th, 1979
Hometown	Pfäfers SG
Marital status	unmarried

Education and Experience

12/2003 – today	City of Zurich Pension Plan, Investment Department
08/2005 – 12/2007	Doctoral Studies Swiss Banking Institute at the University of Zurich
10/1999 – 07/2003	MBA (lic.rer.pol., "magna cum laude") University of Berne
08/1994 – 06/1999	Matur Typus E Städtisches Gymnasium Kirchenfeld Bern,